Geophysical Journal International

Geophys. J. Int. (2015) **202,** 768–786
GJI Geodynamics and tectonics

doi: 10.1093/gji/ggv166

Adjoint-based estimation of plate coupling in a non-linear mantle flow model: theory and examples

Vishagan Ratnaswamy,[1,2] Georg Stadler[3] and Michael Gurnis[2]

[1] *Graduate Aerospace Laboratory of California Institute of Technology (GALCIT), Pasadena, CA* 91125, *USA. E-mail: vratnasw@caltech.edu*
[2] *Seismological Laboratory, California Institute of Technology, Pasadena, CA* 91125, *USA*
[3] *Courant Institute of Mathematical Sciences, New York University, NY* 10012, *USA*

Accepted 2015 April 14. Received 2015 March 31; in original form 2014 December 24

SUMMARY

We develop and validate a systematic approach to infer plate boundary strength and rheological parameters in models of mantle flow from surface velocity observations. Based on a realistic rheological model that includes yielding and strain rate weakening from dislocation creep, we formulate the inverse problem in a Bayesian inference framework. To study the distribution of parameters that are consistent with the observations, we compute the maximum *a posteriori* (MAP) point, Gaussian approximations of the parameter distribution around that MAP point, and employ Markov Chain Monte Carlo (MCMC) sampling methods. The computation of the MAP point and the Gaussian approximation require first and second derivatives of an objective function subject to non-linear Stokes equations; these derivatives are computed efficiently using adjoint Stokes equations. We set up 2-D numerical experiments with many of the elements expected in a global geophysical inversion. This setup incorporates three subduction zones with slab and weak zone (interplate fault) geometry consistent with average seismic characteristics. With these experiments, we demonstrate that when the temperature field is known, we can recover the strength of plate boundaries, the yield stress and strain rate exponent in the upper mantle. When the number of uncertain parameters increases, there are trade-offs between the inferred parameters. These trade-offs depend on how well the observational data represents the surface velocities, and on the weakness of plate boundaries. As the plate boundary coupling drops below a threshold, the uncertainty of the inferred parameters increases due to insensitivity of plate motion to plate coupling. Comparing the trade-offs between inferred rheological parameters found from the Gaussian approximation of the parameter distribution and from MCMC sampling, we conclude that the Gaussian approximation—which is significantly cheaper to compute—is often a good approximation, in particular locally around the MAP point. Thus, the method can be applied to the global problem of inferring non-linear constitutive parameters and plate coupling factors for each subduction zone in a global geophysical inversion with known slab structure.

Key words: Inverse theory; Plate motions; Mantle processes; Rheology: mantle.

1 INTRODUCTION

Plate motion is likely primarily driven by slab pull, the concentration of negative buoyancy from subducted oceanic plates at convergent margins (Forsyth & Uyeda 1975; Chapple & Tullis 1977), while other driving forces, including ridge push and traction at the base of plates, also contribute. Although estimates of slab pull suggest that it may be responsible for more than about 70 per cent of the total driving force (Conrad & Lithgow-Bertelloni 2002), the relative strength of forces controlling plate motion has been difficult to firmly establish because of the diverse origin of driving mechanisms and the close association between slab pull and resisting forces. The slab pull force arises from concentration of negative buoyancy and through thermally activated mantle rheology leads to a larger effective strength of slabs. Slabs both concentrate the driving force [because they act as stress guides (Elsasser 1969)] and concentrate resistance [because the relatively strong slabs bend in the hinge zone (Conrad & Hager 1999)]. There is a significant effect from dynamic weakening at plate boundaries manifested as the development of normal faulting and diffuse seismicity in the outer rise associated with plate bending (Kikuchi & Kanamori 1995). Presumably, as plates are underthrusted at plate boundaries, there is a large stress

concentration within the hinge zone leading to the growth of faults which dissipates energy and weakens the plate. Slabs are a fundamental, highly non-linear component of the system of plate forces, but deconvolving these forces and quantifying uncertainty has been difficult, especially in global models meant to predict and explain actual plate motions.

At convergent plate boundaries, in addition to plate bending, the resistance to plate motion is also likely strongly influenced by sliding of subducting with respect to overriding plates. Such motion gives rise to great earthquakes that occur along many, but not all, subduction zones. Seismic coupling (ratio of seismic moment release to the rate implied by plate motion) varies substantially between subduction zones. Variation in seismic coupling is likely strongly influenced tectonically, that is by local convergence, plate age and the geometry of both the shallow and deep structure of slabs (Ruff & Kanamori 1983). Seismic coupling could also be influenced by the nature of the material [such as the thickness of the sediment on the incoming plate (Ruff 1989)] within the zone between the plates. Such quantities vary substantially not only between subduction zones but also along strike of a plate boundary. A simple heuristic model suggests that highly coupled subduction zones (such as Peru or Chile) may have large normal forces, while weakly coupled subduction zones (such as the Marianas or the Izu-Bonin) have small normal forces (Scholz & Campos 1995, 2012).

The importance of driving and resisting forces has been evaluated with forward and inverse models (Hager & O'Connell 1981; Forte & Peltier 1987; Conrad & Lithgow-Bertelloni 2002). Some spherical models essentially invert observed plate motions for the radial viscosity of the mantle while being driven by the long-wavelength distribution of mantle buoyancy (Forte & Peltier 1987). The models balance stresses from piecewise rigid plates (spherical caps) against the resistance from viscous linear mantle flow with small lateral variations in viscosity.

Such spherical inversions do not include the essential character of slabs which act as stress guides while resisting plate motions through plate bending. By incorporating constitutive relationships with thermally activated diffusion and dislocation creep (Karato & Wu 1993) with yielding, regional 2-D (Billen & Hirth 2005, 2007) and 3-D Cartesian models (Zhong *et al.* 1998) capture the potentially relevant processes within the bending plate while producing plate-like surface motion. Such models require high resolution locally (∼1 km), especially within the hinge zone of the subducting plates, and robust Stokes solvers that can handle the many orders of magnitude variations in effective viscosity implied by laboratory-based constitutive relationships (variations can be six to eight orders of magnitude). The ability to incorporate these resolutions and solvers in spherical models capable of achieving global plate motions has been a computational challenge, and only recently overcome (Stadler *et al.* 2010). Through adaptive mesh refinement, the sharp gradients in viscosity within and near slabs have been achieved in instantaneous models that forward predict global plate motions (Alisic *et al.* 2010, 2012). Such models are also able to show finer scale tectonic motions, such as rapid trench rollback, while not requiring plates to be rigid. As such forward models only approximately matched observed plate motions, their full potential has yet to be achieved.

There has been important progress developing inverse models of mantle flow that bring different sets of data together towards the inference of geophysical properties. For example, adjoints have been implemented with different combinations of the convection equations to infer the initial temperature distribution in the mantle in both regional (Ismail-Zadeh *et al.* 2004; Spasojevic *et al.* 2009) and global (Bunge *et al.* 2003; Horbach *et al.* 2014) contexts. Nevertheless, all of these models use simplified rheologies that do not incorporate the essential physics of strain rate weakening and yielding that are arguably essential for computing the driving and resisting forces of plate motions.

Here, we explore how the high-resolution flow models can be recast as an inverse problem capable of resolving the complex role of slabs and plate margins. We develop a method to infer plate boundary strength, yield stress and strain rate exponent by fitting plate motions in high-resolution models in which slabs and hinge zones are well resolved. We expand on the work of Worthen *et al.* (2014) in three ways. First, the test problems employed here are functionally equivalent to the expected geophysical inverse problem. Secondly, we formulate the problem as a Bayesian inverse problem, which allows a more complete characterization of the physical trade-offs and the uncertainties in the inferred parameters. Thirdly, we provide expressions for the second derivatives (the Hessian matrix) of the mismatch functional between the geophysical observations and model predictions. For that purpose, we derive the first-order adjoint equations along with the expression of the gradient of this misfit functional with respect to the parameters. A Bayesian formulation of the inverse problem allows us to quantify uncertainties in the inferred parameters in addition to computing the best-fit, that is, maximum *a posteriori* (MAP), parameters. The computation of the MAP estimate amounts to solving a PDE-constrained optimization problem, for which we employ an inexact Newton conjugate gradient method. To estimate the uncertainty in these parameters we explore their posterior distribution, that is, the solution of the Bayesian inverse problem. We compare results obtained from sampling the posterior distribution with its Gaussian approximation centred at the MAP parameters. With a series of computations of a 2-D model problem, we demonstrate the trade-offs between the mechanical properties that occur in the system. We demonstrate that the Gaussian approximation of the *a posteriori* distribution is a reasonable approximation of the posterior distribution near the MAP point. We then discuss issues associated with applying the methods to the spherical global mantle flow problem to infer of the parameters in the non-linear constitutive relationship and the spatial distribution of plate coupling from present day plate motions.

2 NON-LINEAR STOKES FORWARD PROBLEM

We model mantle flow with the infinite Prandtl-number Boussinesq approximation, which leads to the following non-dimensional Stokes equations:

$$\nabla \cdot \sigma = -\mathrm{Ra}\, T \mathbf{e}_\mathbf{r} \qquad \text{on } \Omega, \tag{1a}$$

$$\nabla \cdot u = 0 \qquad \text{on } \Omega, \tag{1b}$$

where Ω is the mantle domain (assumed to be 2-D in this paper), $\sigma = \sigma(u, p) = 2\eta\dot{\varepsilon}(u) - p\mathbf{I}$ is the stress tensor with the viscosity $\eta = \eta(\dot{\varepsilon}_{\mathrm{II}}, \Gamma, T, \sigma_y)$, which depends on the velocity u (through the second invariant of the strain tensor $\dot{\varepsilon}_{\mathrm{II}}$ defined below), on multiplicative factor Γ modelling plate boundaries, on the temperature T and on the yield stress $\sigma_y > 0$. Moreover, $\dot{\varepsilon}(u) := \frac{1}{2}(\nabla u + \nabla u^T)$ is the strain rate tensor, p is the pressure and $\mathbf{I}$ the identity tensor. The Stokes eqs (1) are driven by thermal buoyancy. Here, $\mathrm{Ra} = \frac{\rho g \alpha \Delta T D^3}{\kappa \eta_{\mathrm{ref}}}$ is the thermal Rayleigh number, where ρ is the density of the mantle, g is the gravitational acceleration, α is the thermal diffusivity, ΔT is the

temperature difference, D is the length scale, η_{ref} is the reference viscosity and κ is the thermal diffusivity. The second invariant of the strain rate tensor is $\dot{\varepsilon}_{\text{II}} = \frac{1}{2}[\text{tr}(\dot{\varepsilon}^2(u)) - \text{tr}(\dot{\varepsilon}(u))]$. In the limit of incompressibility considered here, $\dot{\varepsilon}_{\text{II}}$ reduces to

$$\dot{\varepsilon}_{\text{II}} = \frac{1}{2}\text{tr}(\dot{\varepsilon}^2(u)). \tag{2}$$

Note that, in the geophysics literature sometimes the square root of $\dot{\varepsilon}_{\text{II}}$ is referred to as the second invariant of the strain rate tensor. No normal flow and free-slip tangential conditions on the boundary $\partial\Omega$ of Ω are used, that is

$$u \cdot \mathbf{n} = 0, \quad \mathbf{T}(\sigma\mathbf{n}) = 0 \qquad \text{on } \partial\Omega \tag{3}$$

Here, we use the tangential operator for the Neumann condition defined as $\mathbf{T} = \mathbf{I} - \mathbf{n}\mathbf{n}^T$ is the projection onto the tangential direction. In particular, plate velocities on the top are not imposed but are an outcome of model calculations.

In the following, we prefer to work with the weak (variational) form of the Stokes eqs (1). This weak form is derived by multiplying (1a) and (1b) by arbitrary functions v and q, respectively, which are assumed to be sufficiently smooth, and satisfy the equivalent Dirichlet boundary condition, $v \cdot \mathbf{n} = 0$. Using integration by parts and the boundary conditions (3), this results in

$$\int_\Omega 2\eta(\dot{\varepsilon}_{\text{II}}, \Gamma, n, \sigma_y)\dot{\varepsilon}(u) : \dot{\varepsilon}(v)\text{d}\Omega - \int_\Omega p\nabla \cdot v \,\text{d}\Omega - \int_\Omega q\nabla \cdot u \,\text{d}\Omega$$
$$= \int_\Omega \text{Ra}T\mathbf{e}_r \cdot v \,\text{d}\Omega. \tag{4}$$

On geological timescales, the mantle behaves like a viscous fluid from thermally activated creep. The viscosity strongly depends on temperature, and this dependence can be represented by an Arrhenius-type law. In the upper mantle, dislocation creep likely dominates over diffusion creep (Stocker & Ashby 1973). Although one can prescribe the rheology as composite (Billen & Hirth 2007; Stadler *et al.* 2010) such that both, diffusion and dislocation creep can play a role depending on the state of stress and the strain rate, we have found that dislocation creep dominates within the plates and slabs and hence is the deformation mechanism which likely controls plate motions. Thus, underlying our models is a temperature-dependent shear-thinning rheology,

$$\tilde{\eta}(\dot{\varepsilon}_{\text{II}}, T) = \Gamma a(T)\dot{\varepsilon}_{\text{II}}^{\frac{1-n}{2n}}, \text{ with } \Gamma(x) = 1 - \sum_i (1 - \Gamma_i)\chi_i(x),$$

where $a(T) := \text{A}_{\text{rad}}\exp(\beta(0.5 - T))$, and $\chi_i(\,\cdot\,)$ are characteristic functions for individual plate boundaries, that is a function with value 1 at the (volumetrically modelled) plate boundary, and a value of 0 away from the plate boundary. The strength/weakness of the coupling along plate boundaries is controlled by the weakening factors $\Gamma_i > 0$. Plate decoupling occurs over long timescales within seismogenic zones, where great earthquakes typically occur. The degree of frictional resistance that occurs along the seismogenic zone is controlled by the factors Γ_i: small values of Γ_i give rise to weakly coupled plate boundaries, while larger values enforce stronger coupling. Plate boundaries require high-spatial resolution in computational models, and the coupling factors Γ_i will act as parameters in the inversion.

An important aspect of the mantle rheology is dynamic weakening through shear thinning, in particular near hinge zones. Thus, we use a rheology that involves plastic yielding additionally to polynomial shear thinning. For computational reasons we also incorporate lower and upper viscosity bounds $0 < \eta_{\min} < \eta_{\max}$ in the rheology, such that the (effective) viscosity is

$$\eta(\dot{\varepsilon}_{\text{II}}, \Gamma, n, \sigma_y) = \eta_{\min}$$
$$+ \min\left(\Gamma \min\left(\eta_{\max}, a(T)(\dot{\varepsilon}_{\text{II}} - d)^{\frac{1}{2n}}\dot{\varepsilon}_{\text{II}}^{-\frac{1}{2}}\right), \frac{1}{2}\sigma_y\dot{\varepsilon}_{\text{II}}^{-\frac{1}{2}}\right). \tag{5}$$

Here, roots of negative quantities are considered to be zero; they do not play a role due to the viscosity bounds. The choice (5) for the effective viscosity corresponds to first applying the upper viscosity bound to the temperature and strain rate dependent viscosity. This is followed by the multiplication with $\Gamma(x)$, a function describing plate boundaries through low viscosity zones. Finally, the plastic yielding condition is imposed. Adding $\eta_{\min}$ enforces a lower bound on the viscosity, as well as a one-to-one correspondence between strain rate and stress in the case of plastic yielding. In (5), we use a shift $d \geq 0$—which is derived in Appendix A—to ensure that $\eta_{\max}$ is incorporated in a way that the viscosity is continuously differentiable with respect to $\dot{\varepsilon}_{\text{II}}$, and thus also with respect to the velocity. This differentiability is important as we target Newton-type methods for the solution of the non-linear Stokes equations and also require derivatives in the inversion. Note however that the continuous differentiability of the viscosity with respect to $\dot{\varepsilon}_{\text{II}}$ does not hold when plastic yielding occurs.

There exists several areas in the mantle where dynamic weakening or the viscosity bounds are important. For instance, the condition $\eta_{\max} < \eta(\dot{\varepsilon}_{\text{II}}, \Gamma, T, \sigma_y)$ primarily holds in the oceanic lithosphere where strain rate weakening plays a secondary role and the viscosity structure is dominated by the temperature. There are two possible cases: (i) the effective viscosity is $\eta_{\max} + \eta_{\min}$, if $\eta_{\max}$ is smaller than the yielding viscosity or, (ii) $\eta_{\text{eff}} = \frac{1}{2}\sigma_y\dot{\varepsilon}_{\text{II}}^{-1/2} + \eta_{\min}$, that is yielding occurs. While $\eta_{\max}$ is dominant in the lithosphere, in the asthenosphere the viscosity is primarily dominated by dislocation creep, that is $\eta(\dot{\varepsilon}_{\text{II}}, \Gamma, T, \sigma_y) < \eta_{\max}$.

An important characterization of the system state, both within subduction zones and the whole model domain, is the viscous dissipation in a subdomain $\tilde{\Omega} \subset \Omega$ of interest given by

$$\Phi(\tilde{\Omega}) = \int_{\tilde{\Omega}} 2\eta(\dot{\varepsilon}_{\text{II}}, \Gamma, n, \sigma_y)\dot{\varepsilon}(u) : \dot{\varepsilon}(u)\,\text{d}\tilde{\Omega}. \tag{6}$$

Viscous dissipation has been analysed in models of subduction zones and related to the bending of plates (Conrad & Hager 1999): as plates bend at subduction zones, a large amount of energy associated with dynamic weakening is released.

3 BAYESIAN INVERSION

A Bayesian approach to inverse problems allows one to infer the most likely parameters together with their uncertainties from a computational model, observational data, and prior knowledge on the parameters. In our problem, the parameters to be inferred are the plate coupling strength coefficients Γ_i and rheological parameters; the observational data are the observed plate velocities, and prior knowledge on the parameters can come, for example, from laboratory experiments or inferences from other geophysical observations. The computational model describes our theory of how parameters and observations are related. In our case, this relation is given through the solution of the non-linear Stokes equations discussed in Section 2. We collect all parameters in a vector $m \in \mathbb{R}^p$, collect the plate velocity observations in a vector u_{obs} and denote the mathematical model that maps parameters to plate velocities by $f(m)$. Note that for given parameters m, the computation of $f(m)$ is computationally costly, as it requires the solution of non-linear Stokes

equations, followed by extracting the corresponding plate velocities from the velocity field data. Additionally, even if the involved Stokes operator were linear, the parameter-to-observable map $f(\cdot)$ is non-linear as the Stokes flow depends non-linearly on rheological parameters.

In our Bayesian inversion approach, we assume that observation and model errors follow a Gaussian distribution with zero mean and covariance matrix $\mathcal{C}_{\text{noise}}$, that is

$$f(m) - u_{\text{obs}} = \mathcal{N}(0, \mathcal{C}_{\text{noise}}).$$

Thus, the likelihood probability density function (pdf), which describes the likelihood of observations u_{obs} for given model parameters m is given by

$$\pi_{\text{like}}(u_{\text{obs}}|m) \propto \exp\left(-\frac{1}{2}(f(m) - u_{\text{obs}})^T \mathcal{C}_{\text{noise}}^{-1}(f(m) - u_{\text{obs}})\right),$$

where '$\propto$' denotes proportionality up to a (normalization) constant that makes π_{like} a proper density. Additionally, we assume a given prior pdf $\pi_{\text{prior}}(m)$, which incorporates our prior knowledge on the parameters. Bayes' theorem states that the posterior probability distribution $\pi_{\text{post}}(m)$, which is the solution of the Bayesian inverse problem, is given by (Kaipio & Somersalo 2005; Tarantola 2005)

$$\pi_{\text{post}}(m) \propto \pi_{\text{like}}(u_{\text{obs}}|m)\pi_{\text{prior}}(m). \tag{7}$$

Even if the prior is Gaussian, say with mean m_0 and covariance matrix $\mathcal{C}_{\text{prior}}$, that is

$$\pi_{\text{prior}}(m) \propto \exp\left(-\frac{1}{2}(m - m_0)^T \mathcal{C}_{\text{prior}}^{-1}(m - m_0)\right), \tag{8}$$

the posterior pdf given by (7) is, in general, non-Gaussian due to the non-linearity of the parameter-to-observable map. Statistical estimators to explore and characterize the posterior pdf π_{post}—and thus the solution of the Bayesian inverse problem—are discussed next.

3.1 Exploring the posterior distribution

Important statistical estimators for a distribution are the *MAP* point, the *mean* and the *covariance matrix*. We next briefly discuss these estimators and their approximations. For simplicity of the presentation, we assume a Gaussian prior given by (8), such that the posterior pdf has the form

$$\pi_{\text{post}}(m) \propto \exp\Big(-\frac{1}{2}(f(m) - u_{\text{obs}})^T \mathcal{C}_{\text{noise}}^{-1}(f(m) - u_{\text{obs}}) - \frac{1}{2}(m - m_0)^T \mathcal{C}_{\text{prior}}^{-1}(m - m_0)\Big). \tag{9}$$

The parameter vector m, where π_{post} takes its maximum is called the MAP point, m_{MAP}. It can be found by maximizing π_{post}, or equivalently, by minimizing the negative log of the posterior pdf, that is

$$\min_{m \in \mathbb{R}^p} \mathcal{J}(m), \tag{10}$$

where

$$\mathcal{J}(m) := \frac{1}{2}(f(m) - u_{\text{obs}})^T \mathcal{C}_{\text{noise}}^{-1}(f(m) - u_{\text{obs}}) + \frac{1}{2}(m - m_0)^T \mathcal{C}_{\text{prior}}^{-1}(m - m_0). \tag{11}$$

This problem has the form of the regularized least squares optimization problem occurring in deterministic inverse problems, with the term coming from the prior corresponding to the regularization used to cope with the ill-posedness common to inverse problems (Vogel 2002; Tarantola 2005). Effective minimization of (11) requires computation of derivatives of $\mathcal{J}$ with respect to the parameters m. The computation of these derivatives is complicated by the fact that the parameter-to-observable map involves the solution of a partial differential equation. We use adjoint equations to make this computation efficient (see Section 3.2). Building on these derivatives, we use a Newton method for the solution of (10): Starting from an initial guess m^0 for the parameters, for $k = 1, \ldots$ one computes a Newton update direction $\bar{m}$ by solving

$$\mathcal{H}(m^k)\bar{m} = -\mathcal{G}(m^k) \tag{12a}$$

and updates

$$m^{k+1} = m^k + \alpha\bar{m}, \tag{12b}$$

where $\mathcal{G}$ and $\mathcal{H}$ denote the gradient and the Hessian with respect to m, respectively. In (12a), $\alpha > 0$ is a step length, which is, starting from an initial step length of $\alpha = 1$ reduced using backtracking to ensure descent of the negative log likelihood $\mathcal{J}(\cdot)$ from the kth to the $(k+1)$st iteration; see Nocedal & Wright (2006).

To explore the posterior pdf beyond the MAP estimate—and thus quantify the uncertainty in the parameter estimates—one can either characterize the posterior through sample statistics or construct an approximation to π_{post}. We use both of these approaches and compare the results and their computational efficiency.

Sampling methods (Hastings 1970; Gilks 2005) allow the complete characterization of π_{post}, but they often require a large number of evaluations of $f(m)$, that is many non-linear Stokes solves for different parameters. Sampling is particularly challenging for high-dimensional parameter vectors m, as the posterior pdf is defined over a space of the dimension of the parameter vector; this difficulty for high-dimensional distributions is often referred to as *curse of dimensionality*.

An alternative to sampling is to construct a Gaussian approximation of π_{post}, centred at the MAP point. For that purpose, we consider the linearization F of the parameter-to-observable map $f(\cdot)$ at the MAP estimate m_{MAP}:

$$f(m) \approx f(m_{\text{MAP}}) + F(m - m_{\text{MAP}}). \tag{13}$$

Using this approximation for the parameter-to-observable map in (9) results—after rearranging terms—in a Gaussian approximation π_{post}^G of the posterior given by

$$\pi_{\text{post}}^G(m) \propto \exp\left(-\frac{1}{2}(m - m_{\text{MAP}})^T \left(F^T \mathcal{C}_{\text{noise}}^{-1} F + \mathcal{C}_{\text{prior}}^{-1}\right)(m - m_{\text{MAP}})\right).$$

Note that the mean of this Gaussian approximation π_{post}^G is m_{MAP}, and the covariance matrix is the inverse of $(F^T \mathcal{C}_{\text{noise}}^{-1} F + \mathcal{C}_{\text{prior}}^{-1})$, which is the Hessian of $\mathcal{J}$ after linearization of the parameter-to-observable map.

The computation of the MAP point and of the Gaussian approximation of the posterior about the MAP point require derivatives of the negative log posterior $\mathcal{J}(\cdot)$ with respect to the parameter vector m. As $\mathcal{J}(\cdot)$ depends on the solution of the non-linear Stokes equations, we use adjoint methods to compute these derivatives efficiently, which is the topic of the next section.

3.2 Adjoint-based computation of derivatives

Adjoint equations allow one to efficiently compute derivatives of scalar-valued functions (such as $\mathcal{J}(\cdot)$) with respect to a large number of parameters by solving a single (linear) adjoint equation. Using finite differences (or forward sensitivities) is an alternative to the use of adjoint methods, but has the disadvantage that it requires the solution of an equation for each parameter, that is each component of m. Hence, for a large number of parameters and expensive-to-solve forward models (as the non-linear Stokes equations (1)), the use of adjoint methods to compute derivates is crucial.

We choose a parameterization for the coupling factors Γ_i, the yield stress σ_y, and the strain rate exponent n that ensures their positivity by considering their logarithms as the inference parameter vector, that is

$$m = (\log(\Gamma_1), \log(\Gamma_2), \ldots, \log(\sigma_y), \log(n))^T. \tag{14}$$

The parameter-to-observable map $f(\cdot)$ maps the parameters m to the Stokes velocity u (or some linear function $\mathcal{O}u$ of the Stokes velocity) on the top boundary $\partial\Omega_t$ of Ω. The flow velocity observations can either be pointwise field observations, or observations of the average velocity of plate-like structures. Furthermore, the operator $\mathcal{O}$ extracts the velocities at points on the surface corresponding to the physical points of the observations. The latter models the type of geophysical observations that are mostly available, in which plates are considered rigid.

Using the form of the parameter-to-observable map to specify the negative log likelihood function (11) in the context of our target problem results in

$$\mathcal{J}(m) = \frac{1}{2}(\mathcal{O}u(m) - u_{\text{obs}})^T \mathcal{C}_{\text{noise}}^{-1}(\mathcal{O}u(m) - u_{\text{obs}}), \tag{15}$$

where $u(m)$ denotes the solution of the non-linear Stokes equations for the parameters m. The function (15) represents the misfit of the observed surface velocities u_{obs} with surface velocities from simulations. For simplicity of the notation, in (15) we neglect the quadratic contribution coming from the prior, which is simple to differentiate.

We use a Lagrangian method (Tröltzsch 2010; Borzì & Schulz 2012) to compute derivatives of $\mathcal{J}$ with respect to the parameters m. For that purpose, we define a Lagrangian functional as the sum of the objective (15), and the weak form of the Stokes eq. (4). In the objective, we consider m and the Stokes velocity u as independent variables and thus write $\mathcal{J}(m, u)$ rather than $\mathcal{J}(m)$. In the weak form of the Stokes eq. (4), the test functions v and q take the role of Lagrange multiplier functions, which satisfy adjoint equations, which will be derived below. We thus refer to the Lagrange multipliers v and q as adjoint velocity and pressure, respectively. The Lagrangian functional is defined as follows:

$$\begin{aligned}\mathcal{L}(u, p, v, q, m) = \mathcal{J}(m, u) + \int_\Omega 2\eta(\dot{\varepsilon}_{\text{II}}, \Gamma, n, \sigma_y)\dot{\varepsilon}(u) : \dot{\varepsilon}(v)\,\mathrm{d}\Omega \\ - \int_\Omega p\nabla\cdot v\,\mathrm{d}\Omega - \int_\Omega q\nabla\cdot u\,\mathrm{d}\Omega - \int_\Omega \mathrm{Ra}T\mathbf{e}_r\cdot v\,\mathrm{d}\Omega.\end{aligned} \tag{16}$$

Then, the gradient $\mathcal{G}(m)$ is given by the gradient of $\mathcal{L}$ with respect to m, provided all variations of the Lagrangian with respect to (u, p) and (v, q) vanish, see (Tröltzsch 2010; Borzì & Schulz 2012). Thus, we next derive expressions for $\mathcal{L}$ with respect to m and derive the equations that must be satisfied if all other variations of the Lagrangian vanish.

Taking variations of the Lagrangian with respect to m results in the following expressions for the ith component of the gradient

$$\mathcal{G}(m)_i = \int_\Omega 2\eta_{,i}(\dot{\varepsilon}_{\text{II}}, \Gamma, n, \sigma_y)\dot{\varepsilon}(u) : \dot{\varepsilon}(v)\,\mathrm{d}\Omega, \tag{17}$$

where we have used that $\eta_{,i}(\dot{\varepsilon}_{\text{II}}, \Gamma, n, \sigma_y)$ denotes the derivative of the viscosity η with respect to the parameter m_i. To compute derivatives of η with respect to m_i, we distinguish how the different parameters enter in the rheology. If $m_i = \log(\Gamma_i)$ is the log of the ith weak factor, this derivative is given by

$$\eta_{,i}(\dot{\varepsilon}_{\text{II}}, \Gamma, n, \sigma_y) = \begin{cases} 0 & \text{in } \Omega_y, \\ \Gamma_i \chi_i \min\left(\eta_{\max}, a(T)(\dot{\varepsilon}_{\text{II}} - d)^{\frac{1}{2n}}\dot{\varepsilon}_{\text{II}}^{-\frac{1}{2}}\right) & \text{in } \Omega \setminus \Omega_y. \end{cases}$$

where $\Gamma_i = \exp(m_i)$. Here, we denote by $\Omega_y \subset \Omega$ the points where yielding occurs, that is where $\eta(\dot{\varepsilon}_{\text{II}}, \Gamma, n, \sigma_y) = \eta_{\min} + 1/2\sigma_y\dot{\varepsilon}_{\text{II}}^{-1/2}$. At these points, the Stokes solution is not sensitive to m_i. Next, we consider the case that $m_i = \log(\sigma_y)$ and we obtain the derivative

$$\eta_{,i}(\dot{\varepsilon}_{\text{II}}, \Gamma, n, \sigma_y) = \begin{cases} \frac{1}{2}\sigma_y\dot{\varepsilon}_{\text{II}}^{-\frac{1}{2}} & \text{in } \Omega_y, \\ 0 & \text{in } \Omega \setminus \Omega_y. \end{cases}$$

Finally, if $m_i = \log(n)$, we obtain

$$\eta_{,i}(\dot{\varepsilon}_{\text{II}}, \Gamma, n, \sigma_y) = \begin{cases} \Gamma a(T)\omega(\dot{\varepsilon}_{\text{II}} - d)^{\frac{1}{2n}}\dot{\varepsilon}_{\text{II}}^{-\frac{1}{2}} & \text{in } \Omega_w, \\ 0 & \text{in } \Omega \setminus \Omega_w, \end{cases}$$

where $\omega = \log((\dot{\varepsilon}_{\text{II}} - d)^{-\frac{1}{2n}})$ and $\Omega_w \subset \Omega$ are the points where $\eta(\dot{\varepsilon}_{\text{II}}, \Gamma, n, \sigma_y) = \eta_{\min} + a(T)(\dot{\varepsilon}_{\text{II}} - d)^{1/(2n)}\dot{\varepsilon}_{\text{II}}^{-1/2}$, and thus the viscosity depends on the strain rate exponent n. If we did not use the parameterization of $\log(m)$, but rather inverted for m, then $\omega = \log((\dot{\varepsilon}_{\text{II}} - d)^{-\frac{1}{2n^2}})$ which is similar to the procedure in Petra *et al.* (2012) and Worthen *et al.* (2014).

Requiring that variations of the Lagrangian with respect to the adjoint velocity and pressure (v, q) vanish is equivalent with the (weak form of) the forward Stokes equations. Setting all variations of the Lagrangian with respect to the forward velocity and pressure (u, p) to zero, and subsequent integration by parts leads to the adjoint equations, which characterize the adjoint velocity v in (17). These adjoint equations are given by

$$\begin{aligned} \nabla\cdot v &= 0 \qquad \text{on } \Omega, \\ \nabla\cdot\hat{\sigma}_u &= 0 \qquad \text{on } \Omega, \end{aligned} \tag{18}$$

with boundary conditions

$$\begin{aligned} v\cdot\mathbf{n} &= 0 \quad \text{on } \partial\Omega, \\ \mathbf{T}(\hat{\sigma}_u\mathbf{n}) &= \begin{cases} 0 & \text{on } \partial\Omega \setminus \partial\Omega_t, \\ -\mathcal{O}^T\mathcal{C}_{\text{noise}}^{-1}(\mathcal{O}u - u_{\text{obs}}) & \text{on } \partial\Omega_t, \end{cases} \end{aligned}$$

where $\hat{\sigma}_u = \hat{\sigma}_u(v, q)$ is the adjoint stress tensor defined by

$$\hat{\sigma}_u = 2\left(\eta(\dot{\varepsilon}_{\text{II}}, \Gamma, n, \sigma_y)\mathbb{I} + \frac{1}{2}\eta_{,\dot{\varepsilon}_{\text{II}}}[\dot{\varepsilon}(u)\otimes\dot{\varepsilon}(u)]\right)\dot{\varepsilon}(v) - q\mathbf{I} \tag{19}$$

with $\mathbb{I}$ being the fourth-order identity tensor, and $\eta_{,\dot{\varepsilon}_{\text{II}}}$ given by

$$\eta_{,\dot{\varepsilon}_{\text{II}}} = \begin{cases} \min\left(0, \Gamma a(T)(\dot{\varepsilon}_{\text{II}} - d)^{\frac{1}{2n}}\dot{\varepsilon}_{\text{II}}^{-\frac{1}{2}}\frac{\dot{\varepsilon}_{\text{II}} - (\dot{\varepsilon}_{\text{II}} - d)n}{\dot{\varepsilon}_{\text{II}}(\dot{\varepsilon}_{\text{II}} - d)n}\right) & \text{in } \Omega \setminus \Omega_y \\ -\frac{1}{2}\sigma_y\dot{\varepsilon}_{\text{II}}^{-\frac{3}{2}} & \text{in } \Omega_y. \end{cases} \tag{20}$$

Distinguishing between these two cases is necessary since the viscosity does not necessarily depend continuously on $\dot{\varepsilon}_{\text{II}}$ at points

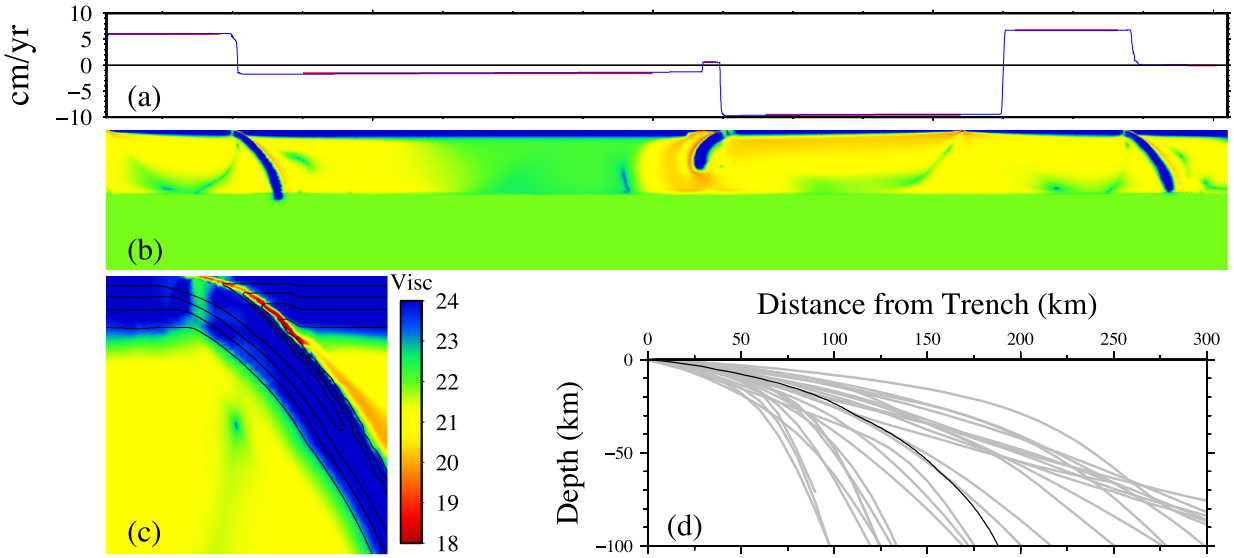

Figure 1. (a) Surface velocity from a forward calculation (thin blue line) and subsampling used to create the data for the inverse models (red lines). (b) $\log_{10}$ of the effective viscosity in Pa s. (c) Detail of the effective viscosity for the left-most subduction zone overlaid by contours of temperature (in black). (d) Profile of the weak zone used in our dynamic model (black line) compared against a representative sample of global seismic coupling zones taken from the Slab 1.0 model (Hayes *et al.* 2012) (in grey).

where the yielding criterion is active. The anisotropic fourth-order tensor in (18) originates from the differentiation of the second invariant with respect to the velocity u. In particular,

$$[\dot{\epsilon}_{\mathrm{II}}]_{,u}(\tilde{u})\,\dot{\varepsilon}(u) = [\dot{\varepsilon}(u) : \dot{\varepsilon}(\tilde{u})]\,\dot{\varepsilon}(u)$$
$$= [\dot{\varepsilon}(u) \otimes \dot{\varepsilon}(u)]\,\dot{\varepsilon}(\tilde{u}),$$

where in the last equality we have used the identity $(a : b)c = (c \otimes b)a$ for second-order tensors a, b, c, where $\otimes$ is the outer product between tensors. Note that the adjoint eq. (18) is linear in its unknowns (v, q), and it is forced by the misfit in plate velocity data on the top surface $\partial\Omega_t$. The momentum equation involves an anisotropic stress tensor (19), which depends on the forward velocity u.

Second derivatives (i.e. Hessians) are needed to compute the Newton update step for finding the MAP point, and for computing Gaussian approximations of the posterior distribution. This second derivative information can either be calculated through finite differences of gradients, or by taking second derivatives of the Lagrangian function (16) with respect to all variables (Tröltzsch 2010; Borzì & Schulz 2012; Petra *et al.* 2012). Both approaches only provide the application of the Hessian to vectors, and each of these Hessian-vector multiplications comes at the cost of two Stokes-like solves: Taking finite differences between gradients requires the computation of gradients for perturbed parameters, amounting to two (one non-linear forward and one linar adjoint) PDE solves. Taking second variations of the Lagrangian results in two linear PDEs (sometimes called incremental equations or second-order adjoint) that must be solved to compute the application of the Hessian matrix to a vector. Such a Hessian-vector application is sufficient if the Newton system (12a) is solved using the conjugate gradient method. Moreover, it also allows the assembly of the Hessian column-by-column (through application on unit vectors) or, if the number of parameters is too large for this approach, the construction of a low-rank approximation of the Hessian (Bui-Thanh *et al.* 2013). The inverse of the resulting (approximation of the) Hessian matrix is then the covariance matrix of a Gaussian approximation of the posterior, as described in Section 3.1.

4 MODEL SETUP AND NUMERICAL SOLUTION

We setup a 2-D Cartesian problem with many of the principal tectonic elements that are thought to be relevant in driving and resisting plate motions. The model has three subducting plates with different amounts of slab penetration into the lower mantle. This variation is intended to span the range of slab penetration and expected coupling that may exist in the global distribution of plates with some slabs only partially penetrating the upper mantle and others fully embedded into the high viscosity lower mantle . Different sensitivities could exist depending on how much of the lower mantle is coupled into plate motions. One of the subduction zones has an overriding oceanic backarc basin (with a small spreading centre) which can result in trench-rollback. In a subduction zone with a back arc basin and rollback, our ability to infer mantle properties might differ from a margin without trench-rollback since the proportion of buoyancy force pulling the plate can differ significantly between these systems. Moreover, the seismic coupling between subduction zones that have trench-rollback, versus those that do not, is central to the discussion on the occurrence of great earthquakes between subduction zones.

The thermal structure of the lithosphere is characterized by a half-space cooling model for the thermal boundary layer. The thermal structure for each slab is computed as follows. First, the top of the slab is identical with the weak zone interface (see below) projected to depth. We then define an initial thermal structure based on the half-space model with depth measured normal to the curving top of the slab. Finally, we diffuse this initial thermal structure for a time proportional to the transit time to arrive at that depth with the velocity of the subducting plate. This method results in thermal structures close to those obtained from the solution of the coupled flow and advection-diffusion problem with non-linear viscosity (e.g. Billen & Hirth 2007). The slabs conserve their buoyancy compared to the incoming plate, but have realistic thermal gradients, for example between the slab interior and the mantle wedge (Fig. 1c). The domain has a width of 12 000 km and depth of 1500 km. The properties that we ascribe to the model are summarized in Table 1.

Table 1. Physical quantities used in our tests.

Parameter	Value
Density (ρ)	3300 kg m^{-3}
Gravity (g)	9.81 m s^{-2}
Coefficient of thermal expansion (α)	2×10^{-5}
Temperature difference ΔT	1400 K
Depth of layer (D)	1500 km
Thermal diffusivity (κ)	10^{-6} m^2 s^{-1}
Reference viscosity (η_{ref})	10^{20} Pa s
Rayleigh number (Ra)	3.06×10^7
Strain rate exponent in upper mantle (n)	3.0

We assume that the effective viscosity has six orders of magnitude variation, 10^{18}–10^{24} Pa s, across the model domain. The minimum effective viscosity ($\eta = 10^{18}$ Pa s) corresponds to η_{min}, while the maximum effective viscosity (η_{max}) is 10^{24} Pa s. The smallest viscosities occur within subducting plate boundaries and mid-ocean ridges, while the largest occur within the interior of oceanic plates. We tie the effective viscosity η_{eff} within the upper mantle below stationary plates to the inferred value of the mantle viscosity from post-glacial rebound studies by adjusting the pre-exponent A_{rad} in the forward model. A discontinuity at 670 km depth is included, where a non-linear viscosity (5) transitions to constant viscosity, that is, a Newtonian rheology.

The characteristic functions $\chi_i(\cdot)$ for the weak zones modelling individual plate boundaries are parametrized with a Gaussian distribution about a centreline. The centreline of the weak zone (and hence the top surface of the initial thermal slab, as described above) is constructed such that it falls within the middle of slab profiles from the Slab 1.0 model (Hayes *et al.* 2012) of nearly all ocean-ocean and ocean-continent subduction zones (Fig. 1d). The weak zones consequently have a shallow dip (approximately 5 degrees) at the surface and represent a significant improvement over the parametrization we have used in the past (Stadler *et al.* 2010).

Our standard case has the following set of parameters: a stress–strain rate exponent, n, of 3, a yield stress, σ_y, of 128.9 MPa and coupling factors, Γ_i, from left to right of 2×10^{-5}, 10^{-5} and 3×10^{-5} so that the plate coupling varies with the middle plate boundary being the most decoupled and the right most plate boundary the most coupled. This model has a realistic strain rate exponent in the upper mantle, with a yield stress that allows plates to be strong while still being able to weaken as they subduct. The different degree of plate coupling at each subduction zone were set to mimic situations expected in a global model.

We discretize the Stokes equations and their adjoints using finite elements on a locally refined mesh of unstructured quadrilaterals. In particular, the mesh is refined around plate boundaries and hinge zones (to ∼5 km resolution) and around the edges of slabs. It consists of 47 360 elements overall. We use Taylor-Hood finite elements (Elman *et al.* 2005), that is, continuous second-order elements for the velocity components, and continuous first-order elements for the pressure. Our implementation is in `MATLAB`[1], and we use `COMSOL v3.5`[2] for meshing and for the assembly of finite element matrices, similar to the model problems in Petra & Stadler (2011).

Newton's method is used to solve the non-linear state eq. (1). We use the fact that the solution of (1) minimizes a viscous energy functional to ensure convergence of the Newton iteration by reducing the size of the update when the a Newton update step fails to reduce the viscous energy (Petra *et al.* 2012; Worthen *et al.* 2014). A commonly used alternative to Newton's method for the solution of non-linear equations is the Picard fixed point method, which, however, often converges much slower than Newton's method. Additionally, because the linearization of the forward problem is self-adjoint, the operator in the adjoint equations is the same operator that arises in the Newton linearization for the forward Stokes problem. This means that any forward non-linear Stokes solver based on a Newton method is already equipped with the operator needed to solve the adjoint Stokes problem. Our implementation uses a direct, factorization-based solver for the linear(ized) Stokes problems. In a large-scale framework, this direct solver must be replaced by an iterative Stokes solver, for instance a preconditioned Krylov method (Elman *et al.* 2005). We terminate the Newton iterations for the solution of the non-linear Stokes equations when the non-linear residual is reduced by four orders of magnitude.

Computing the MAP point amounts to solving the optimization problem (10), which can be written as an optimization problem with PDE constraints given by the nonlinear Stokes eqs (1). Using derivatives computed through adjoint equations (see Section 3.2), we employ the (inexact) Newton method outlined in Section 3.1 for the solution of this PDE-constrained optimization problem. Here, inexactness refers to the fact that the Hessian system (12a) is not solved exactly, but iteratively through a conjugate gradient method (Nocedal & Wright 2006). This approach only requires Hessian-vector applications rather than the assembled Hessian operator. Each Hessian-vector application requires two linear Stokes solves and assembling the Hessian matrix would require two linear Stokes solves for each parameter, which is infeasible for problems with a large number of parameters. In particular for these problems, solving (12a) using the conjugate gradient method can be a significant advantage compared to constructing the Hessian matrix. We use line search to ensure sufficient decent of the optimization functional and terminate the Newton iterations for the MAP point after the gradient has been reduced by four orders of magnitude.

5 FORWARD MODEL

With the distribution of temperature and weak zones as described for the standard case, we find that the resulting effective viscosity is characterized by strong plates and a weak asthenosphere (Fig. 1). Each of the three subducting oceanic plates move with velocities of 3–10 cm yr^{-1} and with the right most overriding plate being nearly stationary while the other two overriding plates roll-back with velocities of 1–2 cm yr^{-1}. The plates are strong away from plate boundaries and have piecewise constant surface velocities. Below the intersection of the faults (shear zones) with the surface, where the plate starts to bend within the hinge zone, the effective viscosity is reduced as these areas exceed the yield stress. The yield stress and prefactor to the viscosity law (for the given temperature field) were chosen to give these velocity and effective viscosity outcomes, which are similar to those found previously (Zhong *et al.* 1998; Billen & Hirth 2005, 2007)

Plate motions are sensitive to the strength of the plate margins; by plate margin strength we refer to the combination of the strength of the oceanic lithosphere and slab and the strength of the coupling (shear zone) between overriding and subducting plates. As plate margins become progressively weaker, plate motion eventually becomes insensitive to resistive forces at plate boundaries (e.g. King & Hager 1990). This can be seen in the limit when an individual coupling factor, Γ_i, or their average, $\langle\Gamma\rangle$, becomes small

[1] http://www.mathworks.com
[2] http://www.comsol.com

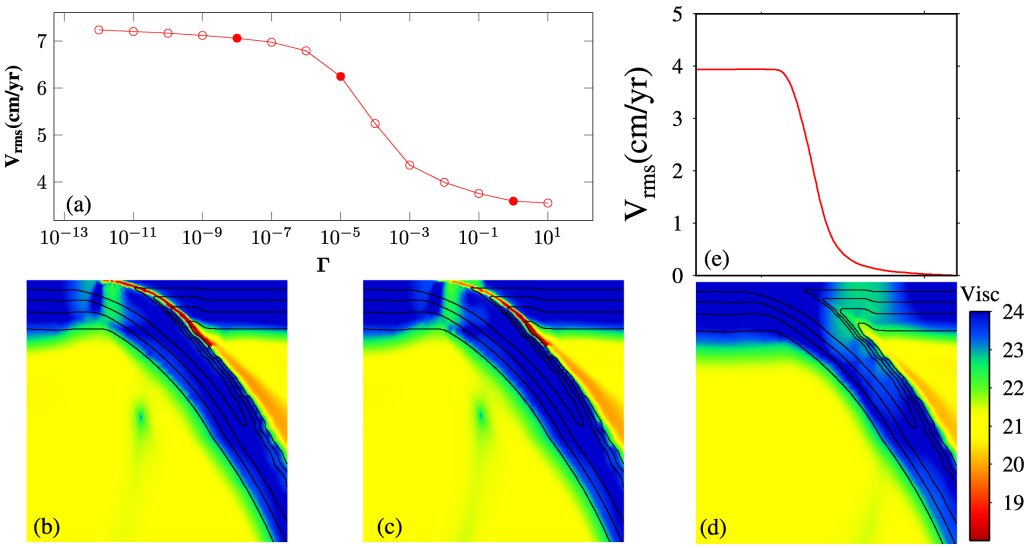

Figure 2. (a) Change in plate velocity (as rms across the top surface) as a function of plate coupling $\langle\Gamma\rangle$ for a set of models in which all three plate margins have the same coupling factor. Filled symbols denote the cases shown in panels b–d. Viscosity structure in the vicinity of right most subduction zone for three values of coupling: (b) $\langle\Gamma\rangle = 10^{-8}$, (c) $\langle\Gamma\rangle = 10^{-5}$ and (d) $\langle\Gamma\rangle = 1$ (e) Surface velocity profile of right most subduction zone at $\langle\Gamma\rangle = 1$.

(Fig. 2a). Here and in the following, $\langle\Gamma\rangle$ is the arithmetic mean of the individual weak zone prefactors Γ_i. As the fault (shear zone) weakens ($\Gamma_i \to 0$) plates are free to slide by each other with the resistance coming from the effective viscosity of the slab and bending plate. As the coupling factor of a plate boundary approaches 10^{-10}, plate boundaries become very weak (Fig. 2); when the value approaches unity, plates become locked and the rms of the surface velocity asymptotes to approximately 2.5 cm yr^{-1}. We will not further investigate the limit of fully locked plates. For our subsequent considerations, we refer to the region $10^{-8} < \langle\Gamma\rangle < 10^{-4}$ to be in the *sensitive* regime because variation in $\langle\Gamma\rangle$ leads to a substantial change in plate velocity, while the region $\langle\Gamma\rangle < 10^{-8}$ will be referred to as the *insensitive* regime. We choose $\langle\Gamma\rangle < 10^{-8}$ to be the insensitive region because the effective viscosity, η_{eff}, approaches the minimum effective viscosity limit. Note that the minimum viscosity η_{min}, however, is only attained exactly for $\Gamma_i = 0$ due to the form of the viscosity given in (5). In Fig. 2(a), as $\langle\Gamma\rangle < 10^{-6}$, the rms plate velocity approaches an asymptotic value slightly larger than 7 cm yr^{-1}. Note that the effective viscosity is bounded from below by $\eta_{\text{min}} = 10^{-2}$; since the characteristic weak zone functions χ_i are they attain their minimum only at the centre and choosing $\Gamma_i < 10^{-6}$ further weakens the plate coupling and thus effects the plate velocity.

As $\langle\Gamma\rangle$ increases, deformation within the overriding plate increases as seen from the effective viscosity in the vicinity of the subduction zones (Figs 2b–d). When the shear zone becomes locked, the deformation shifts from a combination of the bending plate and shear zone to spatially distributed deformation in the overriding plate. The distributed deformation within the overriding plate now occurs over a larger length scale. For a given average viscosity of the plates, we expect that there should be a trade-off when inferring n, Γ, and σ_y from plate motions, as each of these parameters influence plate motions.

6 INVERSE MODEL

In this and the next section, we study the extent to which we can reconstruct the parameters, and find their trade-offs and uncertainties from plate motion data. We use the surface velocities from forward models as synthetic data. To generate this synthetic data, we solve forward problems with known rheological parameters and plate coupling factors, and compute the resulting plate motions. To obtain plate velocity data similar to what is available from a global kinematic model, we generate synthetic data from the computed surface velocities by taking a single scalar average of the velocities away from plate margins. To be precise, we use overall 78 per cent of the top surface to compute rms values for six plates defined in red, amounting to a 5-D data vector (see the thick red lines in Fig. 1a). We do not add noise to these synthetic averaged measurements. The retained data is similar to what is available from a global kinematic model of plate motions, namely Euler poles and associated uncertainties (determined from fracture zone, magnetic lineation, seismic focal mechanism and GPS data) for each plate. The single scalar value for each plate is the 1-D equivalent (over a 2-D mantle cross section) to an Euler pole. These data are indicative of plates that are rigid away from plate margins, but the margins are free to deform if the inferred rheological parameters allow them to. For subducting oceanic plates, we set the areas without surface velocity constraints to be much larger than the expected zone of yielding and bending such that the data do not impose this length scale. We assume uncorrelated observations with a standard deviation of 0.21 cm yr^{-1}, that is, $\mathcal{C}_{\text{noise}}$ is the unit matrix scaled by 0.21^{-2}. The experiments in this section do not incorporate a prior, and thus the MAP point is the maximum likelihood point; for convenience, we still refer to it as the MAP point.

We apply the methods developed above to infer the MAP estimates for the weakening factors Γ_i, the yield stress σ_y and the strain rate exponent n. We also report the standard deviation σ (corresponding to the $\log(m)$) for the Gaussian approximation of the posterior at the MAP point, that is, the diagonal entries in the posterior covariance matrix. In our experiments we vary not only the underlying properties of the mantle flow system, but also which quantities are considered known and thus kept fixed, and which are considered uncertain and inverted for (Table 2).

As initial guess for the computation of the MAP point we choose the weak factors Γ_i equal to 10^{-3}, which is, for most cases, larger than the actual values. This would be a reasonable starting condition

Table 2. Parameter estimation results for several test cases. Each entry for the prefactors, yield stress and strain rate exponent are given as inferred/true/standard deviation (log(m)). Quantities that are considered known and thus fixed in the inversion are shown in parentheses. The top surface velocity data are constant for each plate, and the percentage of the top surface with these constant velocity constraints is indicated in the second column.

Case	V_{rms} (surf.cov./plate)	Γ_1 (left)	Γ_2 (middle)	Γ_3 (right)	σ_y (MPa)	n
I	78 per cent	$2 \times 10^{-5}/2 \times 10^{-5}/0.470$	$10^{-5}/10^{-5}/0.1630$	$3 \times 10^{-5}/3 \times 10^{-5}/0.608$	(128.9)	(3.0)
II	78 per cent	$2 \times 10^{-8}/2 \times 10^{-8}/4.65$	$10^{-8}/10^{-8}/2.35$	$3 \times 10^{-8}/3 \times 10^{-8}/6.36$	(128.9)	(3.0)
III	78 per cent	$2 \times 10^{-5}/2 \times 10^{-5}/0.66$	$10^{-5}/10^{-5}/0.185$	$3 \times 10^{-5}/3 \times 10^{-5}/0.867$	128.9/128.9/0.0969	(3.0)
IV	78 per cent	$2 \times 10^{-8}/2 \times 10^{-8}/6.19$	$10^{-8}/10^{-8}/2.72$	$3 \times 10^{-8}/3 \times 10^{-8}/9.66$	128.9/128.9/0.0833	(3.0)
V	78 per cent	$2 \times 10^{-5}/2 \times 10^{-5}/0.553$	$10^{-5}/10^{-5}/0.310$	$3 \times 10^{-5}/3 \times 10^{-5}/0.0.610$	(128.9)	$3.0/3.0/2.67 \times 10^{-3}$
VI	78 per cent	$2 \times 10^{-8}/2 \times 10^{-8}/5.94$	$10^{-8}/10^{-8}/9.24$	$3 \times 10^{-8}/3 \times 10^{-8}/6.51$	(128.9)	$3.0/3.0/4.31 \times 10^{-3}$
VII	78 per cent	(2×10^{-5})	(10^{-5})	(3×10^{-5})	$128.9/128.9/8.05 \times 10^{-2}$	$3.0/3.0/1.83 \times 10^{-3}$
VIII	78 per cent	(2×10^{-8})	(10^{-8})	(3×10^{-8})	$128.9/128.9/5.37 \times 10^{-2}$	$3.0/3.0/1.1 \times 10^{-3}$
IX	78 per cent	$2.0 \times 10^{-5}/2 \times 10^{-5}/0.564$	$10^{-5}/10^{-5}/0.485$	$3.0 \times 10^{-5}/3 \times 10^{-5}/1.40$	128.9/128.9/0.338	$3.0/3.0/9.11 \times 10^{-3}$
X	78 per cent	$2 \times 10^{-8}/2 \times 10^{-8}/3.29$	$10^{-8}/10^{-8}/5.16$	$3 \times 10^{-8}/3 \times 10^{-8}/5.72$	$128.9/128.9/9.34 \times 10^{-2}$	$3.0/3.0/3.12 \times 10^{-3}$
XI	AllV	$2.0 \times 10^{-5}/2 \times 10^{-5}/0.926$	$10^{-5}/10^{-5}/0.429$	$3.0 \times 10^{-5}/3 \times 10^{-5}/1.37$	128.9/128.9/0.321	$3.0/3.0/8.5 \times 10^{-3}$
XII	78 per cent	$2 \times 10^{-6}/2 \times 10^{-6}/1.64$	$10^{-5}/10^{-5}/0.998$	$3 \times 10^{-5}/3 \times 10^{-5}/1.97$	128.9/128.9/0.471	$3.0/3.0/1.41 \times 10^{-2}$
XIII	AllV	$2 \times 10^{-6}/2 \times 10^{-6}/0.926$	$10^{-5}/10^{-5}/0.429$	$3 \times 10^{-5}/3 \times 10^{-5}/1.37$	128.9/128.9/0.321	$3.0/3.0/8.50 \times 10^{-3}$
XIV	78 per cent/47 per cent (right)	$1.965 \times 10^{-5}/2 \times 10^{-5}/0.40$	$10^{-5}/10^{-5}/0.70$	$2.793 \times 10^{-3}/3 \times 10^{-3}/0.839$	129.9/128.9/0.155	$3.0/3.0/5.27 \times 10^{-3}$
XV	78 per cent/60 per cent (right)	$1.30 \times 10^{-5}/2 \times 10^{-5}/0.404$	$1.07 \times 10^{-5}/10^{-5}/0.680$	$1.26 \times 10^{-3}/3 \times 10^{-3}/0.748$	154/128.9/0.156	$3.017/3.0/5.36 \times 10^{-2}$
XVI	78 per cent/73.3 per cent (right)	$3.32 \times 10^{-6}/2 \times 10^{-5}/0.454$	$1.37 \times 10^{-5}/10^{-5}/0.838$	$5.5 \times 10^{-4}/3 \times 10^{-3}/1.0$	266.1/128.9/0.788	$3.08/3.0/2.77 \times 10^{-2}$
XVII	78 per cent/86.7 per cent (right)	$1.59 \times 10^{-6}/2 \times 10^{-5}/0.592$	$1.71 \times 10^{-5}/10^{-5}/0.857$	$4.6 \times 10^{-4}/3 \times 10^{-3}/0.645$	517/128.9/0.769	$3.22/3.0/3.26 \times 10^{-2}$
XVIII	78 per cent	$10^{-5}/10^{-5}/0.308$	$10^{-5}/10^{-5}/0.308$	$10^{-5}/10^{-5}/0.308$	128.9/128.9/7.34 $\times 10^{-2}$	$3.0/3.0/3.01 \times 10^{-3}$
XIX	78 per cent	$2.79 \times 10^{-8}/2 \times 10^{-6}/9.4$	$1.26 \times 10^{-7}/10^{-5}/12.6$	$10^{-7}/3 \times 10^{-5}/6.91$	$66.5/53.3/2.52 \times 10^{-2}$	$2.99/3.0/8.93 \times 10^{-3}$

for a geophysical inversion as it makes no assumption about which margins are strong or weak, nor about the relative strength of plate coupling. We initialize the strain rate exponent with $n = 2.7$ and the yield stress with $\sigma_y = 84.4$ MPa ($\sigma_y = 190$ MPa in case XIX). We observe in our numerical experiments that these initializations do not influence the recovered parameters, that is, the optimization problem for the MAP point converges to the same solution independent of the initialization. This suggests that the negative log likelihood/posterior $\mathcal{J}(\cdot)$ does not have multiple local minima. For problems with two parameters, this is also suggested by the contour lines of the posterior pdfs shown later, as local minima in $\mathcal{J}(\cdot)$ would correspond to local maxima in the corresponding pdfs. We terminate the optimization problem when the norm of the gradient $\mathcal{G}$ has been decreased by a factor of 10^4.

In the first experiment (Table 2, case I), we assume that the two global quantities, yield stress and strain rate exponent, are known, and we attempt to infer the plate boundary coupling factors, Γ_i, $i = 1, 2, 3$, for each weak zone in the sensitive regime. We recover the pre-factors, and, as a consequence, the shear stresses within each plate boundary within 1 per cent error after seven iterations. We correctly infer the different plate coupling strengths of each plate margin. This is important as the coupling strength for each subduction zone in a global geophysical inversion is expected to be different.

While the prefactors for case I are in the sensitive regime, in case II they are substantially smaller and thus lie in the insensitive regime. Again, the inferred MAP estimate coincides with the true prefactors. However, there is a substantial increase in the uncertainty of the recovered coupling factors compared to case I (Table 2). In Fig. 3, we study the interplay between standard deviation and the coupling factors systematically. We perform experiments that are identical to cases I and II, that is, with fixed strain rate exponent and yield stress, but choose identical true coupling factors $\Gamma_1 = \Gamma_2 = \Gamma_3$ for each model calculation. We find that uncertainties are largest when plate couplings are small or when the plate margin is fully coupled (Fig. 3a), that is, if the plate coupling factors are in the insensitive regime.

In the previous inversions, the yield stress and strain rate exponent were considered known. As the yield stress and strain rate exponent are also uncertain, we attempt to infer each one of them individually along with the prefactors in cases III (inference of Γ_i, σ_y) and V (inference of Γ_i, n). In cases III and V, we correctly infer the respective values. Although the uncertainty increases when the additional parameter is inferred, the computational cost remains approximately the same as we are able to determine the correct value within 1 per cent after seven iterations. Similarly, the correct values are inferred when $\langle\Gamma\rangle$ is in the insensitive region for the otherwise identical cases IV and VI, although the uncertainty on the prefactors does increase within the insensitive compared to the sensitive regime.

We next attempt to infer the global constitutive parameters—the yield stress and strain rate exponent in the upper mantle—while assuming that the individual prefactors, Γ_i, are known for each plate boundary. Although inferring the global strain rate exponent and yield stress is not a realistic geophysical inversion as it assumes knowledge of the prefactors *a priori*, the case is illustrative of the trade-offs likely to be seen in the full inversion of Γ, σ_y and n. We are able to infer the correct strain rate exponent and yield stress for each case after 8–9 iterations, both in the sensitive (case VII) and insensitive (case VIII) regimes. There is no significant difference in the rate of convergence during the inversion of either n or σ_y when $\langle\Gamma\rangle$ is in the sensitive or insensitive regime.

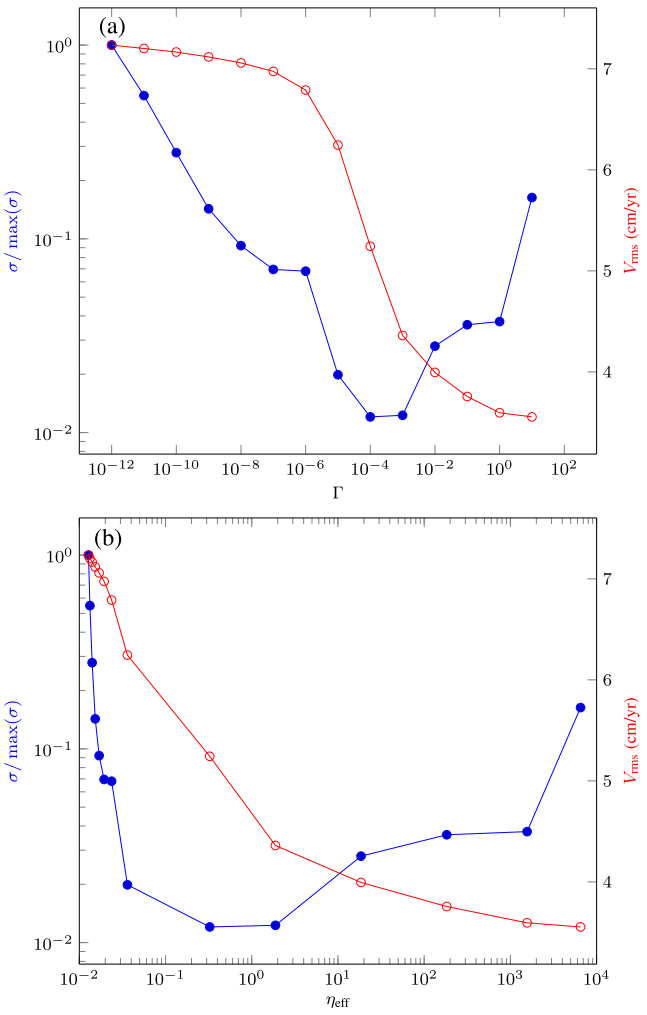

Figure 3. Relationship between the normalized standard deviation σ (in blue) and the average plate velocity $V_{\rm rms}$ (in red) as a function of the plate coupling factor $\Gamma(= \Gamma_1 = \Gamma_2 = \Gamma_3)$ (a), and as a function of the average effective viscosity $\eta_{\rm eff}$ within the weak zones (b).

We now consider a realistic situation where the plate boundary strengths, the yield stress and the strain rate exponent are all unknown, that is, we infer five parameters. Cases IX and X represent inversions where we do not assume any of the rheological parameters in the constitutive relationship other than the activation energy. Despite the potential for trade-offs in the rheological parameters, we are able to recover the true values even when all five parameters are uncertain. However, the uncertainty of the recovered prefactors and rheological parameters increases substantially compared to the earlier cases with fewer parameters left unconstrained. While the values of the parameters are recovered within 1 per cent, it does take more forward–adjoint iterations to converge to the true value: Approximately 10 iterations in the five parameter estimation cases (IX and X) versus 5–6 iterations in the comparable three parameter cases (I, II). We attribute the larger computational cost to the larger number of parameters, and the resulting interplay between the rheological parameters where trade-offs in plate coupling, strain rate weakening and yielding act to minimize the plate velocity misfit. These trade-offs will be explored in the Bayesian inference context in the subsequent section.

With the cases above we showed that parameters (plate couplings, yield stress and strain rate exponent) of an unconstrained system can be inferred when the surface data are approximated with a single rms value for each plate away from plate boundaries. However, there is an important untested assumption of the role played by approximating the surface velocities by a single rms value for each plate. The influence of how the piecewise rms velocity data is applied is demonstrated with additional calculations, in which we change the parts of the top surface that are assumed to move rigidly, and increase a coupling factor such that the plate motion causes deformation in the adjacent plate interior.

From case IX with all of the standard parameters free and 78 per cent of the surface covered with piecewise constant rms values, we compare this against a calculation using all of the surface data without the piecewise constant assumption (XI). Here again, all of the parameters are recoverable except that their variance has been reduced by about a half (Table 2). As the influence and trade-off that occurs in response to surface constraints can be better seen with a larger spread of the coupling factors, we consider cases XII and XIII with the left most coupling factor reduced by 10. Again, all of the parameters are recoverable either with all surface velocities used as data or with 78 per cent as rms values within plate interiors (Figs 4a–c). The variances are again reduced when using the more complete data. Now we retain the greater spread in coupling factors and strain rate exponent but decrease the yield stress from 129 to 53 MPa (case XIX), such that a larger fraction of the lithosphere yields. The plates become more deformable and the average strain rate within the plate interiors increases from 2.22×10^{-16}s^{-1} (case I) to 6.77×10^{-16} s^{-1} (case XIX). Using only the rms values over 78 per cent of each plate, there is a strong trade-off between parameters and only the global strain rate exponent is correctly recovered (Fig. 4e). In this case, the inversion responds to the imposed rigid plate motion data by recovering a larger yield stress (Fig. 4f). The larger yield stress allows the plates to stiffen, but in order to fit the overall magnitude of plate velocities, the MAP estimate has weaker plate margins (Fig. 4d).

The influence of the spatial extent of the imposed data is demonstrated with cases XIV to XVII which are otherwise identical to case IX, except that the coupling factor for the right most subduction zone is increased from the standard value of 3×10^{-5} to 3×10^{-3} (Table 2). In this case, the margin broadens over a length-scale defined by the vertical projection of the weak zone to the surface (Fig. 5a). The rms velocities are now determined over different fractions of the rightmost plate in this series of calculations. When 60 per cent of the surface of the overriding plate is used (Fig. 5c), the MAP point 'responds' by creating a more rigid plate with less yielding immediately below the extend of the imposed data, but adjacent to a zone of deformation (Fig. 5d). Since part of the motion of the subducting oceanic plate is being accommodated by deformation of the overriding plate, the inversion estimates a coupling factor that is smaller than the factor used to generate the synthetic data. If the fraction of the imposed data is enlarged to now encompass nearly the entire surface area of the overriding plate, we infer an entirely rigid overriding plate with a much weaker plate margin (Fig. 5e–f). The recovered yield stress is substantially larger than the actual value so as to decrease the yield stress within the overriding plate. That decreased yield stress nearly eliminates the yielding with the hinge zone. The four cases show that the recovered values progressively deviate when an otherwise deformable plate is forced to be rigid in the inversion (Table 2). In general, the coupling factors decrease and the strain-rate exponent increases. However, when the area of the right most plate with the imposed rms values is reduced to 47 per cent (case XIV), such that there is little rigid plate data applied within the deforming region, the prefactors for

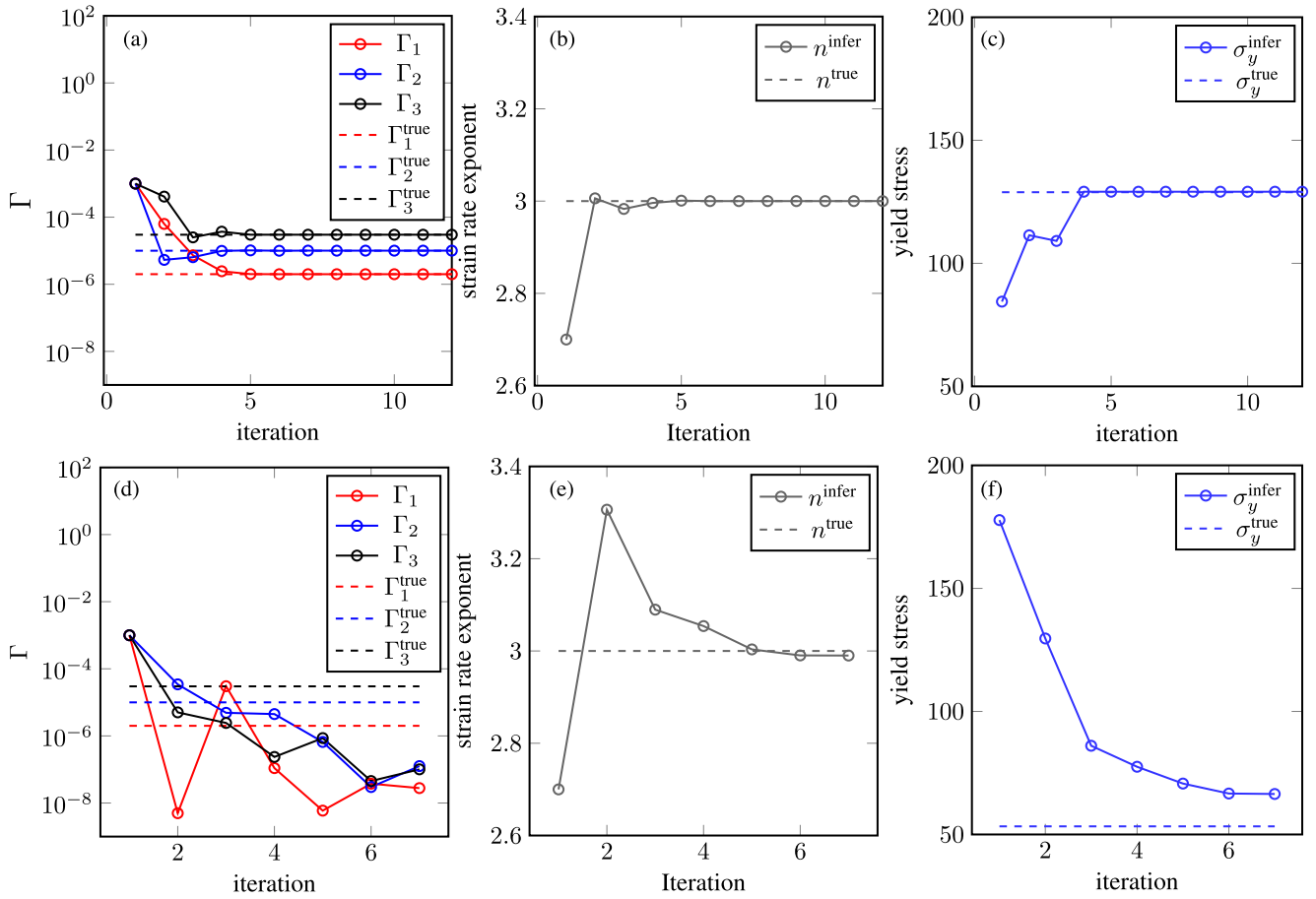

Figure 4. Convergence towards MAP points for case XII (a–c) and for case XIX (d–f). The plots show the convergence of Γ_i, n and σ_y as function of the iteration. The open circles depict the iterates and the dashed horizontal lines show the values used to generate the synthetic data.

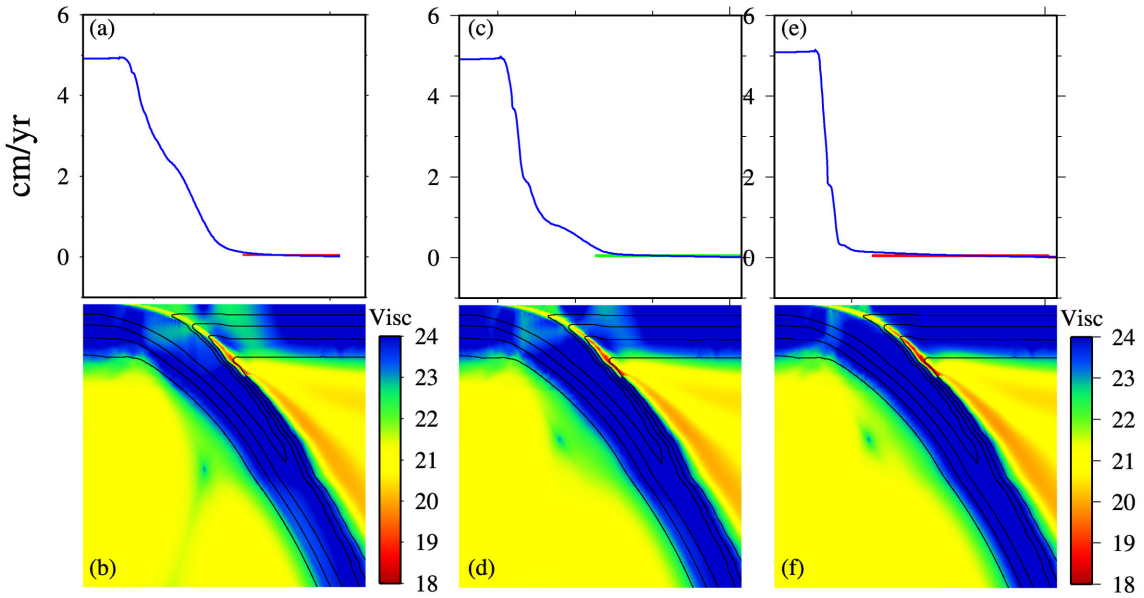

Figure 5. (a) Plate velocities from a forward calculation (blue curve) with a large coupling factor for right-most subduction zone. The distribution of velocities used to compute rms velocities over 87 per cent (pink line) and 47 per cent (green line) over the right most plate. (b) Effective viscosity from the forward model. (c) Inverse with plate velocities (blue) of right-most subduction zone with green line denoting data. (d) Effective viscosity (case XV). (e) Inverse with plate velocities (blue) and pink line denoting data over 87 per cent of right most plate (case XVII). (f) Effective viscosity (case XVII).

Table 3. Dissipation occurring locally and throughout the whole domain for selected cases. Values are shown as recovered/true. Symbols defined in text.

Dissipation	Case IX	Case XIV	Case XVII	Case XIX
$\Phi(\Omega_1^y)$	$9.30 \times 10^7/9.30 \times 10^7$	$1.03 \times 10^8/1.03 \times 10^8$	$7.71 \times 10^7/1.03 \times 10^8$	$9.16 \times 10^8/3.16 \times 10^9$
$\Phi(\Omega_2^y)$	$5.32 \times 10^8/5.32 \times 10^8$	$5.86 \times 10^8/5.86 \times 10^8$	$7.72 \times 10^8/5.86 \times 10^8$	$4.17 \times 10^8/1.67 \times 10^8$
$\Phi(\Omega_3^y)$	$1.46 \times 10^8/1.46 \times 10^8$	$1.12 \times 10^8/1.12 \times 10^8$	$9.18 \times 10^7/1.12 \times 10^8$	$6.77 \times 10^8/1.02 \times 10^9$
$\Phi(\Omega_1^w)$	$4.85 \times 10^7/4.85 \times 10^7$	$5.48 \times 10^7/5.50 \times 10^7$	$8.59 \times 10^7/5.50 \times 10^7$	$4.53 \times 10^7/2.81 \times 10^7$
$\Phi(\Omega_2^w)$	$9.20 \times 10^7/9.20 \times 10^7$	$7.63 \times 10^7/7.65 \times 10^7$	$2.84 \times 10^8/7.65 \times 10^7$	$4.57 \times 10^6/3.35 \times 10^6$
$\Phi(\Omega_3^w)$	$4.57 \times 10^7/4.57 \times 10^7$	$5.99 \times 10^7/4.47 \times 10^7$	$1.26 \times 10^8/4.47 \times 10^7$	$3.14 \times 10^7/1.12 \times 10^7$
$\Phi(\Omega)$	$4.16 \times 10^9/4.16 \times 10^9$	$4.15 \times 10^9/4.18 \times 10^9$	$4.15 \times 10^9/4.69 \times 10^9$	$4.70 \times 10^9/5.80 \times 10^9$

all three subduction zones (as well as the strain rate exponent and yield stress) can be recovered within about 1 per cent, effectively reducing the trade-offs between the inferred parameters.

Finally, we turn to the recovery of dissipation, Φ, (6), a measure of energy dissipated by deformation of a viscous material. As plates accumulate large stresses at subduction zones, there is dynamic weakening near the hinge zones and a concentration of dissipation as oceanic plates subduct into the mantle (Buffett & Rowley 2006; Stadler *et al.* 2010; Alisic *et al.* 2012; Buffett & Becker 2012). To better understand our ability to invert for dissipation, we compute the total dissipation in the whole domain and within the hinge and weak zones (Table 3). Dissipation was determined in those parts of the domain where yielding occurs (mostly in the hinge zones) in the left, middle and right subduction zone and are denoted by Ω_l^y, Ω_m^y, Ω_r^y, respectively; for some extreme models given below, the nodes that yield can differ between forward and recovered models, such that these regions can likewise differ between forward and inverse. For comparison, we also determined the dissipation within low viscosity weak zones Ω_1^w, Ω_2^w, Ω_3^w, respectively for the left, middle and right plate boundaries. We start with case IX, where all of the standard parameters were left unconstrained and all were recovered on inversion. Since all of the velocities and viscosities were essentially identical between forward and the MAP point from the inversion, so too is the dissipation, both locally and through the whole domain (Table 3).

Recovery of parameters was degraded when deforming plates were approximated on recovery with piecewise constant rms velocity data; likewise, recovery of the dissipation was degraded in these cases. When piecewise constant rms velocity data are used over nearly the entire deforming rightmost plate (in response to the larger coupling factor in case XVII, Fig. 5e–f), the dissipation is not well recovered locally. The recovery of the dissipation within the adjacent weak zone is particularly poor and, counter intuitively, the dissipation is overestimated by a factor of two, despite the viscosity within the weak zone being much lower on recovery (Table 2). However, the overriding plate is more rigid and the deformation (and hence dissipation) is shifted from within the plate to the weak zone between the plates. Although, the dissipation is locally not well recovered in some plate boundaries, the total dissipation is only underestimated by 11 per cent. We consider another extreme model with a lower yield stress and hence more deformable plate interiors (case XIX). On inversion, with constant velocity data added to each plate interior, we found trade-offs between the strength of the coupling factors and the yield stress (Fig. 5d–f). Here, we find that the dissipation of each plate boundary is recovered only within a factor of two, while the total dissipation is underestimated by nearly 40 per cent. These problems can be largely avoided if we do not approximate deforming plates with constant velocity data. For example, partial recovery was achieved in case XIV (Table 3) in which the deforming right-most plate was left mostly unconstrained, except in the far-field or several hundred km from the plate edge where the plate acts rigidly (Fig. 5). In this case, there is a slight trade-off in the inferred values (smaller coupling on the right most plate boundary and an increase in yield stress), leading to a larger viscous dissipation in the hinge zone. However, all other measures of dissipation were reasonable well recovered.

7 QUANTIFICATION OF UNCERTAINTY

Inferred parameters are uncertain due to noise in the surface velocity data, trade-offs between rheological parameters, and modelling errors caused by the fact that the mathematical model is an idealized description of the real world. In this section, we explore these uncertainties and trade-offs systematically, and go beyond the Gaussian approximation of the posterior distribution and use sampling to better characterize the posterior distribution.

As inference in a global geophysical system will likely require a large number of parameters while also involving expensive-to-evaluate parameter-to-observable maps, such a full sample-based characterization of the posterior distribution might not be feasible. A Gaussian approximation of the posterior, however, is often computationally tractable, as it usually only requires a number of PDE solves that is comparable to that needed for the computation of the MAP estimate. Naturally, the question arises how well the Gaussian distribution approximates the true posterior distribution. We will study this issue in our 2-D non-linear Stokes flow problem with plates by comparing the Gaussian approximation—which is based on linearization of the parameter-to-observable map about the MAP point—to the true distribution, which is, in general, not Gaussian. To study this true distribution, we either use a regular parameter space grid, or Markov Chain Monte Carlo (MCMC) sampling (Gilks 2005). For this comparison, we use a problem in which we infer three model parameters and compare 2-D conditional and marginal distributions. We interpret the distributions and study the trade-offs between parameters physically.

The three parameters considered in this study are the strain rate exponent n, the yield stress σ_y and the strength of plate coupling Γ (we choose identical plate coupling factors, that is, $\Gamma = \Gamma_1 = \Gamma_2 = \Gamma_3$). In some of our experiments we do not use prior knowledge for these parameters, in others we use prior distributions to incorporate likely ranges or ranges of interest for these parameters. For instance, strain rate exponents $n > 4$ are unlikely (Karato & Wu 1993), as are coupling factors $\Gamma > 10^{-1}$, which would prevent realistic plate motion. Priors can also be used to express the fact that we are not particularly interested in certain parameter regimes, for instance in coupling factors $\Gamma < 10^{-8}$. Based on these considerations, we choose independent Gaussian distributions for the parameters given by $\pi_\Gamma^{\text{prior}} = \mathcal{N}(10^{-5.5}, 3.43)$, $\pi_{\sigma_y}^{\text{prior}} = \mathcal{N}(150\,\text{MPa}, 0.21)$, $\pi_n^{\text{prior}} = \mathcal{N}(2.98, 0.0247)$, where σ_1, σ_2 and σ_3 are the standard deviations for the priors.

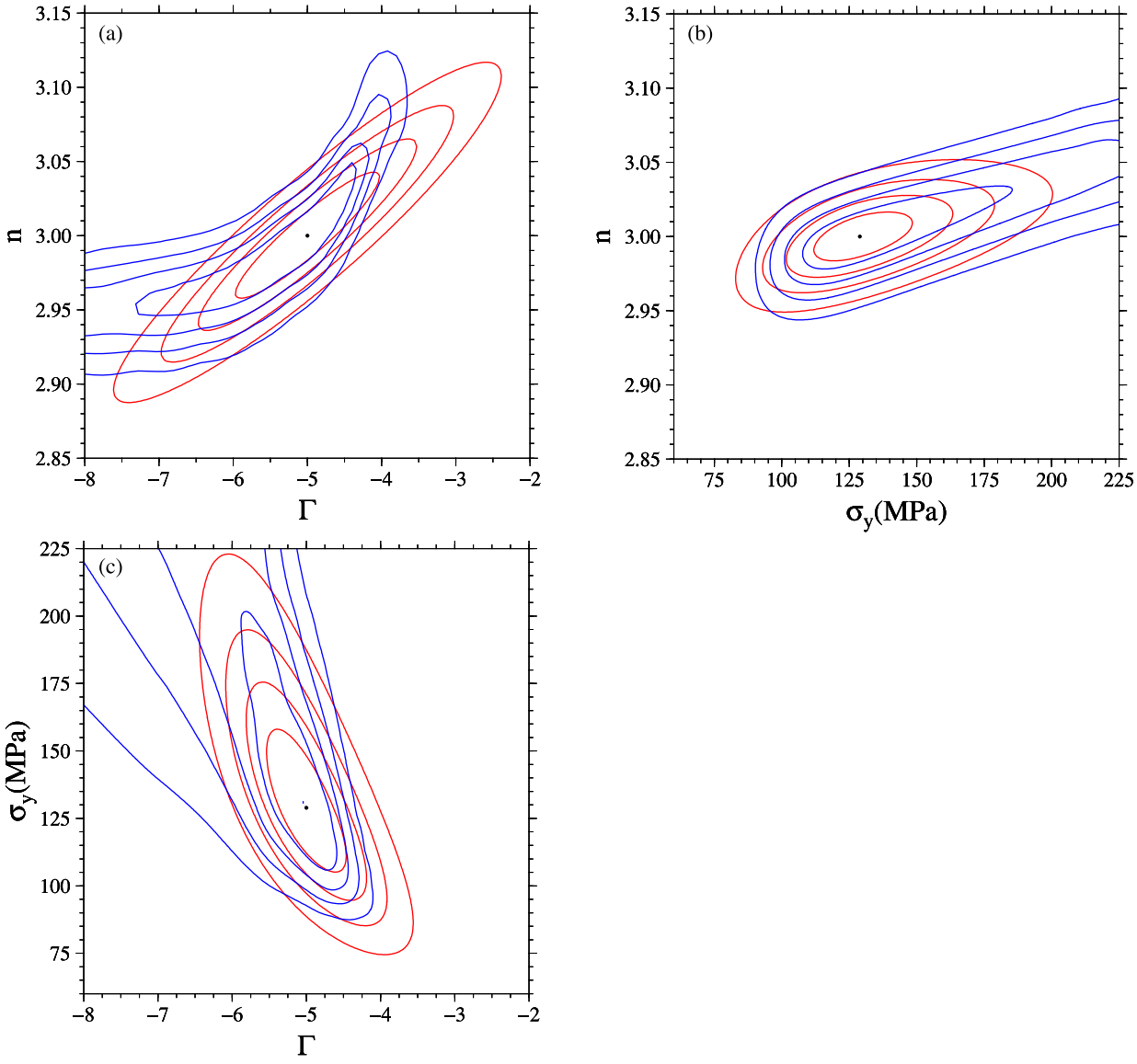

Figure 6. Comparison of 2-D conditional distributions for the three parameters n, Γ and σ_y. Contour lines (for 22, 44, 66, 88 per cent) corresponding to the Gaussian approximation at the MAP point for a setting without a prior are shown in red and for the true posterior distribution are shown in blue. For these conditionals, the parameters kept fixed are in (a) $\sigma_y = 128.9$ MPa, in (b) $\Gamma = 10^{-5}$ and in (c) $n = 3$. Note that due to the parametrization in (14), the contour lines of the Gaussian approximation do not appear as ellipses when plotted in the original n, Γ and σ_y system.

First, we study the 2-D conditional distributions shown in Figs 6 (no prior for parameters) and 7 (with Gaussian prior). Here, we have used a uniformly spaced grid to explore the true posterior distribution, which is clearly not feasible in higher parameter dimensions. In 6(a), we observe a strong trade-off between the magnitude of n and Γ. Models that fit the data with stronger plate coupling require a larger amount of strain weakening, giving an overall positive correlation between n and Γ. The distribution computed from the Hessian (red contours) approximates the true distribution (blue contours) within about one half of a standard deviation. However, as the plate margins become more strongly coupled, an even larger amount of weakening (larger n) than predicted by the Hessian is required, and causes the distribution to bend upward (Fig. 6a). In other words, the model has a non-linearity not accounted for by the linear assumption in (13). The calculation near $\Gamma = 10^{-4}$ and about $n = 3.1$ fits the surface velocity within ≈8 per cent but there is now more distributed deformation within both the overriding and subducting plates. At the other end of the distribution, as the plate margins become weaker, plate motions become less sensitive to plate coupling. As parameters move from the sensitive to the insensitive regime as described previously (Fig. 2a), the best fitting models show a decrease in the strain rate exponent. The conditional distribution 'flattens out' as the slope of the contours become horizontal—as Γ becomes small there are only small changes in the effective viscosity within the weak zone and hence similar data fits are achieved for different Γ (but the same n). This result is partly anticipated with both the insensitivity to plate velocity (Fig. 2a) and the increase in uncertainty (Fig. 3a) with small Γ.

Within the strain rate exponent-yield stress space, we find a conditional distribution from the Hessian near the MAP point that is locally a good approximation to the true conditional distribution (Fig. 6b). The yield stress trades-off with the strain rate exponent with a positive correlation as an increase in yield stress requires an increase in strain rate exponent so as to maintain plate velocities.

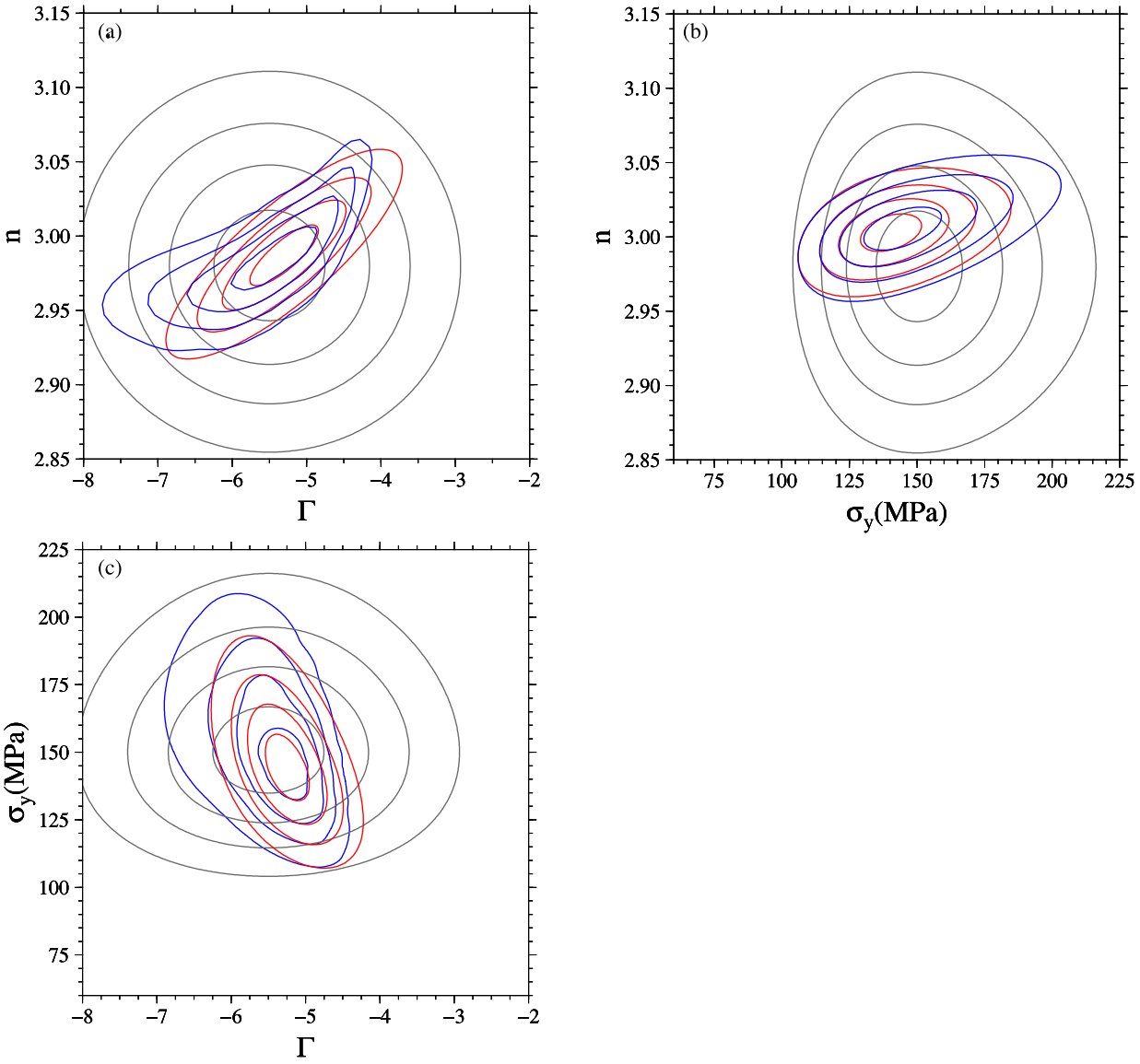

Figure 7. Comparison of 2-D conditional distributions (with Gaussian priors) for the uncertain parameters n, Γ and σ_y. Shown are contour lines (22 per cent, 44 per cent, 66 per cent, 88 per cent) corresponding to the Gaussian approximation at the MAP point (in red), contour lines for the actual posterior distribtion (in blue) and contour lines for the prior distributions (grey). For these conditionals, the parameters kept fixed are in (a) $\sigma_y = 128.9$, in (b) $\Gamma = 10^{-5}$, and in (c) $n = 3$.

However, the slope on the contours of n with respect to σ_y eventually flatten as there is no yielding when σ_y becomes too large. Within the space of yield stress and coupling factors, the conditional shows that the actual distribution is well predicted from the Hessian (Fig. 6c). We find a negative correlation between the prefactor and the yield stress because as the coupling factor between plates increase, the plates need more yielding so as to fit the surface velocity data. The slopes of the contours become constant for small coupling factors when the yield stresses exceed the stresses in the system.

Finally, in Fig. 8, we show 2-D marginals for the Gaussian approximation at the MAP point and compare with marginals of the true posterior distribution. The true distribution is explored using MCMC sampling, and in particular the Delayed Rejection Adaptive Metropolis (DRAM) method (Haario *et al.* 2006). We use 1177 samples computed through repeated forward solves. In Fig. 9, we plot the sample history for the prefactor Γ, which suggests that there is sufficient mixing. The integrated autocorrelation time for the chain τ given by

$$\tau = 1 + 2\sum_{i=1}^{\infty} \rho_k, \tag{21}$$

where $\rho_k = \frac{\mathrm{Cov}[X_t, X_{t+k}]}{\mathrm{Var}[X_t]}$ is the autocorrelation at lag k, with X_t denoting the value of an observed state at time t. The integrated autocorrelation times for each parameters are $\tau_\Gamma = 7.46$, $\tau_{\sigma_y} = 4.68$, $\tau_n = 8.18$. The autocorrelation provides an estimate of the statistical dependence of the samples in the chain. It indicates that about every 5–8th sample in our chain is statistically independent (Robert & Casella 2004). The autocorrelation times should be small (as they are here) so that there are large mean squared jumps, indicating effective mixing and a well sampled posterior distribution.

Next, we qualitatively compare the contours of the 2-D marginals of the Gaussian approximation and the posterior distributions

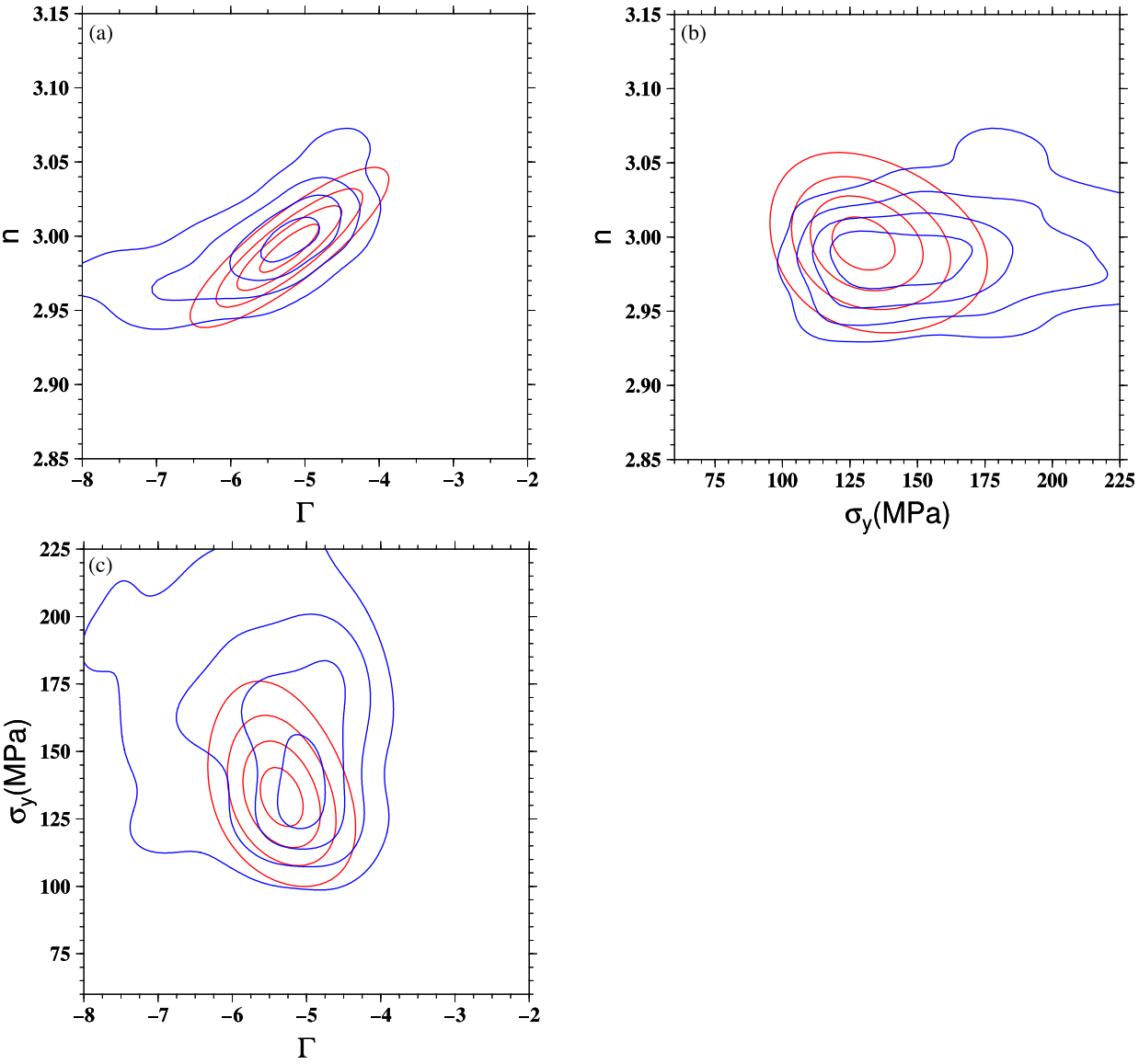

Figure 8. Comparison of 2-D marginal distributions for the unknown parameters n, Γ and σ_y. Contour lines (for 22 per cent, 44 per cent, 66 per cent, 88 per cent) for the marginals from the Gaussian approximation of the posterior distribution are shown in red. Contour lines for the marginals of the true postrior distribution, obtained from an MCMC sampling approach are shown in blue. (a) Marginals distributions for Γ versus n, (b) marginal distributions for σ_y versus n and (c) marginal distributions for Γ versus σ_y.

(Fig. 8). Note that the approximation is reasonable, since the most important trade-offs and correlations found in the posterior and its Gaussian approximation coincide. Compared to the conditionals, however, the difference between the Gaussian approximation of the posterior distribution and the posterior distribution is more pronounced. As to be expected, the marginal of the MAP point does not always coincide with the most likely point of the 2-D marginal. Like the conditionals, the marginals show a positive correlation between Γ and n (Fig. 8a). In all cases, the shifts in the actual distributions away from the Gaussian distributions are caused by (the lack of) yielding for large values of σ_y.

8 DISCUSSION AND CONCLUSIONS

In model problems, we have shown that non-linear constitutive parameters and individual coupling factors between subducting and overriding plates can be inferred along with estimates of uncertainty and the trade-offs between them. Although idealized, the forward models are functionally equivalent to existing highly resolved (1 km where needed) global models (Alisic *et al.* 2010, 2012; Stadler *et al.* 2010), such that the methods developed here will be applicable to parameter inference with quantified uncertainties for the global mantle flow and plate motion problem. Our primary goal here is to discuss present results in terms of their applicability to the geophysical problem.

We use adjoint variables to efficiently compute first and second derivatives of the negative log likelihood function. This requires the solution of the adjoint Stokes eqs (18), which have an anisotropic viscosity but are linear in the adjoint variables. Due to the self-adjointness of the Stokes equations, this adjoint operator coincides with the linear operator required in the Newton method. Hence, a forward non-linear Stokes solver for (1) based on a Newton method is already equipped with the operator needed to solve the adjoint Stokes problem and only the computation of the adjoint system

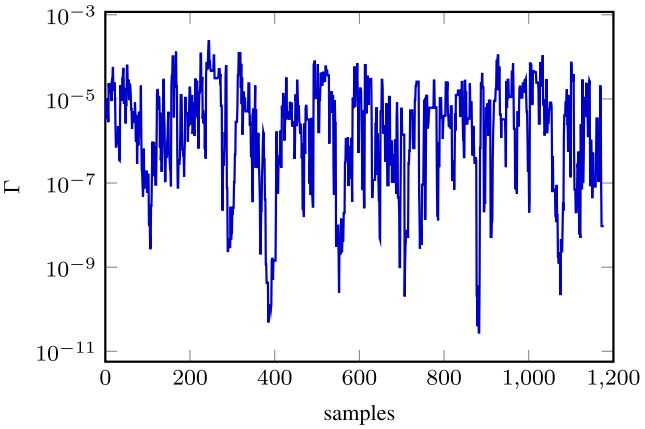

Figure 9. Chain history for weak zone factor Γ.

right hand side must be implemented additionally. Note that the computation of the gradient using adjoints requires a *single* linear (adjoint) solve *independently* of the number of parameters.

As an alternative to computing derivatives through adjoints, finite differentiating for the parameters can be used. In the present problem, this amounts to solving a *non-linear* Stokes equation *for each parameter* to compute the gradient. While for the small number (at most five) of parameters considered in this paper this finite difference approach is certainly practicable, it becomes infeasible for problems with a larger number of parameters, or for problems where a (discretized) parameter field is inferred. Here, we have employed finite differences to verify the implementation of the adjoint-based derivatives.

We described two regimes of subduction-driven plate motion, a sensitive and an insensitive regime, determined from the change in plate motion with respect to coupling factors (e.g. Fig. 2), a relationship known for some time (King & Hager 1990). The distinction between these regimes became evident through individual inversions where we found that the uncertainty was minimized in the sensitive regime and increased with both larger and smaller coupling factors (in the insensitive regime). This suggests that in a global inversion, like in the 2-D test problem, recovery of coupling factors between plate pairs within the sensitive regime should be better determined than factors at plate margins which are either fully uncoupled or fully coupled. It must be emphasized that there will be global interaction between the coupling factors.

The MAP point accurately recovers the coupling factors for each plate boundary in problems where we either assume that the parameters acting over the entire domain (a strain rate exponent and yield stress) are known or when these parameters are inferred simultaneously. The question arising for the global geophysical inversion is whether coupling factors can be inferred in a relative sense (for example, Chile is five times more coupled than the Marianas) or an absolute sense (for example, Chile has a stress of 100 MPa while the value for the Marianas is less than 20 MPa). The method looks promising in both regards. In cases when the plates are not yielding away from plate boundaries, the relative order of plate coupling and their absolute values are recoverable (e.g. Figs 4a–c). We found that the inferred MAP values were insensitive to how we choose the initial guess for the inversion, that is by using larger or smaller constant guesses than the actual values lead to the correct inference of the relative degree of coupling and absolute values of plate coupling. If prior knowledge on either the strain rate exponent or yield stress are added—either by eliminating the parameters from the inversion or by using a smaller variance on the prior—the coupling factors are better constrained. In a global inversion, we could find that the use of a single, constant yield stress might be limiting and disguises the variability in the degree of yielding between subduction zones, perhaps reflected in variable strengths inferred for the bending oceanic lithosphere (Arredondo & Billen 2012). In other words, variability that actually occurs from one bending plate to another (due say to different plate strengths) could be mapped into coupling coefficients. This should be a small effect in a global inversion that uses detailed prior constraints on the thickness of plates, as incorporated into present forward models (Stadler *et al.* 2010; Alisic *et al.* 2010, 2012).

Moreover, as only an instantaneous Stokes flow model is used, this approach could hide the possibility that the development of weakening within a bending plate is a cumulative and not an instantaneous phenomenon. There is nothing inherent in the adjoint-based inference approach that enforces the recovery of only a single, global strain rate exponent n, but our suspicion is that we do not yet have data constraints to sufficiently constrain variations in n. The strain rate exponent, as well as other constitutive parameters may also be variable due to putative variations in major element composition or water content.

How the surface velocity data are used as constraints has an important influence on how close the inverted MAP estimate is to the true parameters, and also influences the trade-offs between parameters. If the horizontal velocity field along the entire surface is used, a complete recovery is possible for the globally acting parameters (yield stress and strain rate exponent) and local parameters (coupling factors). This finding is consistent with the results in Worthen *et al.* (2014) who attempted to recover spatially variable parameters in a non-linear constitutive relationship. Although in this study and in Worthen *et al.* (2014) only the lateral component(s) of the velocity vector along the top are used, this surface velocity field contains powerful information on the absolute value of plate motions and how rapidly and over what length scales plates are deforming. This is particularly useful to constrain the internal deformation of the bending plate. Unfortunately, such pointwise velocity data does not uniformly exist globally, and where is does, for example over continental margins (from dense networks of continuous GPS-stations), the velocity vectors reflect a combination of co- and post-seismic transients and long-term deformation over and adjacent to plate boundaries (McCaffrey *et al.* 2013; Wells & McCaffrey 2013). Below, we discuss how such data could be used in a geophysical inversion. Our new study here deviates from Worthen *et al.* (2014) in that we use limited and piecewise constant plate velocity data, making this study closer to a global problem in which mostly rigid plate motion data are available.

Available plate motion data have passed through a plate motion inversion (Argus & Gordon 1991) such that there is no strain within the plates. Our inversion needs to be sufficiently flexible so as not to impose the length scale or degree of deformation near plate margins, which we have attempted by limiting (in our standard case) to constant velocities from data only over about 80 per cent of the surface within plate interiors. We find that by restricting an inversion to only this data, we are able to recover all of the uncertain parameters to nearly the same degree as when we use the full set of data (for example, compare cases IX and XI in Table 2). This high degree of recovery was achieved when the plates were essentially rigid within their interiors, which is generally a good assumption for most plates. If the plates are not rigid, then we find a strong trade-off between the coupling factors and the parameters governing the non-linearity over the entire model domain. Determining the correct stencil, for example the area over which plate motion data

are prescribed, will be important because some plates, especially the Indian Plate, have present-day internal, but diffuse, deformation (Gordon & Stein 1992; Gordon *et al.* 1998). Based on the 2-D test problems, if we assume that the entire Indian plate were rigid, then we suspect that we would infer incorrect coupling factors or yield stress.

More extreme levels of plate deformation occur closer to plate margins usually within continental areas, such as the Himalayas, Andes, and western North America. In the latter region, plate deformation is particularly well constrained from continuous GPS data, and shows a mixture of distributed deformation associated with plate divergence, transcurrent motion and plate convergence (Kreemer *et al.* 2012). In some 2-D test inverse models, we approximated a plate margin, otherwise deforming over a length scale of several hundred kilometers, with a single constant velocity (akin to a single Euler pole). In such cases, a strong trade-off was found between the inferred coupling factors, both for the adjacent margin and more distant margins, and the degree of yielding through the entire domain. There are likely to be at least two ways of addressing this trade-off in a geophysical inference. First, one could avoid constraining the deforming areas entirely by using a tailored stencil that avoids the deforming areas. Here we should be able to recover the yield stress and strain rate exponent as well as the coupling factors for most plate margins. Another direction would be to attempt an inverse model which combines the recovery of the discrete parameters corresponding to the rigid motion of the plate interiors, as we have done here, with a recovery of a continuous field, like effective viscosity (Worthen *et al.* 2014). This latter method would benefit from the use of distributed velocity constraints. It may now be possible to use such data for western North America using the results of studies which have attempted to deconvolve the short term cycling associated with co- and post-seismic phenomena from long term deformation (McCaffrey *et al.* 2013; Wells & McCaffrey 2013).

In global inversions, the ratio between the number of parameters and the number of independent observations will influence the degree of ill-posedness of the inversion, and it will influence the importance of incorporating prior knowledge for the parameters. The observational data will consist of Euler poles for both major plates, like the Pacific Plate, and minor plates, such as those that make up back-arc basins; consequently we would expect about 12–20 mostly independent observations in a global inversion that was like the test cases described here. Additionally, one could incorporate topography of oceanic trenches, essentially regionally distributed data, that would add to the amount of available observations. This, however, would require some modification to (18) and (19). The uncertain parameters will primarily be the coupling factors for each of the major subductions zones, as well as a small number of globally defined constitutive parameters, which could add up to overall about 20–30 inversion parameters. Note that the non-linearity of the parameter-to-observable map makes it difficult to use the number of parameters and observations directly to judge if the problem is over- or underdetermined—this is the case even if the prior and noise covariance operators are diagonal.

By comparing conditional and marginal distributions, we have observed that Gaussian approximations of the posterior parameter distribution based on the Hessian of the negative log posterior, are useful approximations of the true posteriors, which we explored using MCMC sampling. For global geodynamics inversions, each model evaluation requires significant computational resources and, thus, sampling-based methods, which can require tens of thousands of forward solves even for a moderate number of parameters (Baumann *et al.* 2014), are out of the question. For these problems, a Hessian-based approximation that uses adjoints to compute derivatives is an attractive option to study trade-offs and the interplay between parameters. In particular, Gaussian approximations to conditional and marginal distributions will be important for inversions in global, highly resolved mantle flow models, for instance in the inversion of plate coupling between different subduction zones, which can provide an explanation of how one subduction zone influences another one.

ACKNOWLEDGEMENTS

Supported by the National Science Foundation through EAR-1247022, EAR-1118239, and CMMI-1028978. We thank Noemi Petra and Johann Rudi for valuable discussions.

REFERENCES

Alisic, L., Gurnis, M., Stadler, G., Burstedde, C., Wilcox, L.C. & Ghattas, O., 2010. Slab stress and strain rate as constraints on global mantle flow, *Geophys. Res. Lett.,* **37**(22), doi:10.1029/2010GL045312.

Alisic, L., Gurnis, M., Stadler, G., Burstedde, C. & Ghattas, O., 2012. Multi-scale dynamics and rheology of mantle flow with plates, *J. geophys. Res.: Solid Earth,* **117**(B10), doi:10.1029/2012JB009234.

Argus, D. & Gordon, R., 1991. No-net-rotation model of current plate velocities incorporating plate motion model NUVEL-1, *Geophys. Res. Lett.,* **18,** 2039–2042.

Arredondo, K.M. & Billen, M.I., 2012. Rapid weakening of subducting plates from trench-parallel estimates of flexural rigidity, *Phys. Earth planet. Int.,* **196–197,** 1–13.

Baumann, T.S., Kaus, B.J.P. & Popov, A.A., 2014. Constraining effective rheology through parallel joint geodynamic inversion, *Tectonophysics,* **631,** 197–211.

Billen, M.I. & Hirth, G., 2005. Newtonian versus non-Newtonian upper mantle viscosity: Implications for subduction initiation, *Geophys. Res. Lett.,* **32**(19), doi:10.1029/2005GL023457.

Billen, M.I. & Hirth, G., 2007. Rheologic controls on slab dynamics, *Geochem. Geophys. Geosyst.,* **8**(8), doi:10.1029/2007GC001597.

Borzì, A. & Schulz, V., 2012. *Computational Optimization of Systems Governed by Partial Differential Equations,* SIAM.

Buffett, B.A. & Becker, T.W., 2012. Bending stress and dissipation in subducted lithosphere, *J. geophys. Res.: Solid Earth,* **117**(B5), doi:10.1029/2012JB009205.

Buffett, B.A. & Rowley, D.B., 2006. Plate bending at subduction zones: Consequences for the direction of plate motions, *Earth planet. Sci. Lett.,* **245**(1), 359–364.

Bui-Thanh, T., Ghattas, O., Martin, J. & Stadler, G., 2013. A computational framework for infinite-dimensional Bayesian inverse problems Part I. The linearized case, with application to global seismic inversion, *SIAM J. Sci. Comput.,* **35**(6), A2494–A2523.

Bunge, H.-P., Hagelberg, C.R. & Travis, B.J., 2003. Mantle circulation models with variational data assimilation: inferring past mantle flow and structure from plate motion histories and seismic tomography, *Geophys. J. Int.,* **152**(2), 280–301.

Chapple, W.M. & Tullis, T.E., 1977. Evaluation of the forces that drive the plates, *J. geophys. Res.,* **82**(14), 1967–1984.

Conrad, C.P. & Hager, B.H., 1999. Effects of plate bending and fault strength at subduction zones on plate dynamics, *J. geophys. Res.: Solid Earth,* **104**(B8), 17 551–17 571.

Conrad, C.P. & Lithgow-Bertelloni, C., 2002. How mantle slabs drive plate tectonics, *Science,* **298**(5591), 207–209.

Elman, H.C., Silvester, D.J. & Wathen, A.J., 2005. *Finite Elements and Fast Iterative Solvers with Applications in Incompressible Fluid Dynamics,* Oxford University Press.

Elsasser, W.M., 1969. Convection and stress propagation in the upper mantle, in *The Application of Modern Physics to the Earth and Planetary Interiors,* pp. 1–41, ed. Runcorn, S.K., John Wiley and Sons.

Forsyth, D. & Uyeda, S., 1975. On the relative importance of the driving forces of plate motion, *Geophys. J. Int.,* **43**(1), 163–200.
Forte, A.M. & Peltier, W.R., 1987. Plate tectonics and aspherical earth structure: the importance of poloidal-toroidal coupling, *J. geophys. Res.,* **92**(B5), 3645–3679.
Gilks, W.R., 2005. *Markov Chain Monte Carlo,* Wiley Online Library.
Gordon, R. & Stein, S., 1992. Global tectonics and space geodesy, *Science,* **256,** 333–342.
Gordon, R.G., DeMets, C. & Royer, J.-Y., 1998. Evidence for long-term diffuse deformation of the lithosphere of the equatorial Indian Ocean, *Nature,* **395,** 370–374.
Haario, H., Laine, M., Mira, A. & Saksman, E., 2006. DRAM: efficient adaptive MCMC, *Stat. Comput.,* **16**(4), 339–354.
Hager, B.H. & O'Connell, R.J., 1981. A simple global model of plate dynamics and mantle convection, *J. geophys. Res.: Solid Earth (1978–2012),* **86**(B6), 4843–4867.
Hastings, W.K., 1970. Monte carlo sampling methods using Markov chains and their applications, *Biometrika,* **57**(1), 97–109.
Hayes, G.P., Wald, D.J. & Johnson, R.L., 2012. Slab 1.0: a three-dimensional model of global subduction zone geometries, *J. geophys. Res.: Solid Earth,* **117,** B01302, doi:10.1029/2011JB008524.
Horbach, A., Bunge, H.-P. & Oeser, J., 2014. The adjoint method in geodynamics: derivation from a general operator formulation and application to the initial condition problem in a high resolution mantle circulation model, *Int. J. Geomath.,* **5,** 163–194.
Ismail-Zadeh, A., Schubert, G., Tsepelev, I. & Korotkii, A., 2004. Inverse problem of thermal convection: numerical approach and application to mantle plume restoration, *Phys. Earth planet. Int.,* **145**(1–4), 99–114.
Kaipio, J. & Somersalo, E., 2005. *Statistical and Computational Inverse Problems,* Vol. 160 of Applied Mathematical Sciences, Springer-Verlag.
Karato, S. & Wu, P., 1993. Rheology of the upper mantle: A synthesis, *Science,* **260**(5109), 771–778.
Kikuchi, M. & Kanamori, H., 1995. The Shikotan earthquake of October 4, 1994: lithospheric earthquake, *Geophys. Res. Lett.,* **22**(9), 1025–1028.
King, S.D. & Hager, B.H., 1990. The relationship between plate velocity and trench viscosity in Newtonian and power-law subduction calculations, *Geophys. Res. Lett.,* **17**(13), 2409–2412.
Kreemer, C.W., Hammond, C., Blewitt, G., Holland, A.A. & Bennett, R.A., 2012. A geodetic strain rate model for the Pacific-North American plate boundary, western United States, *Nevada Bureau of Mines and Geology Map,* **178,** 6785.
McCaffrey, R., King, R.W., Payne, S.J. & Lancaster, M., 2013. Active tectonics of northwestern U.S. inferred from GPS-derived surface velocities, *J. geophys. Res.: Solid Earth,* **118,** 1–15.
Nocedal, J. & Wright, S.J., 2006. *Numerical Optimization,* 2nd edn, Springer-Verlag.
Petra, N. & Stadler, G., 2011. *Model variational inverse problems governed by partial differential equations, Tech. Rep. 11-05,* The Institute for Computational Engineering and Sciences, The University of Texas at Austin.
Petra, N., Zhu, H., Stadler, G., Hughes, T. J.R. & Ghattas, O., 2012. An inexact Gauss-Newton method for inversion of basal sliding and rheology parameters in a nonlinear Stokes ice sheet model, *J. Glaciol.,* **58**(211), 889–903.
Robert, C.P. & Casella, G., 2004. *Monte Carlo Statistical Methods,* Vol. 319, Citeseer.
Ruff, L.J., 1989. Do trench sediments affect great earthquake occurrence in subduction zones?, *Subduction Zones Part II,* pp. 263–282, Springer-Verlag.
Ruff, L. & Kanamori, H., 1983. Seismic coupling and uncoupling at subduction zones, *Tectonophysics,* **99**(2), 99–117.
Scholz, C.H. & Campos, J., 1995. On the mechanism of seismic decoupling and back arc spreading at subduction zones, *J. geophys. Res.,* **100**(B11), 22 103–22 115.
Scholz, C.H. & Campos, J., 2012. The seismic coupling of subduction zones revisited, *J. geophys. Res.: Solid Earth (1978–2012),* **117**(B5), doi:10.1029/2011JB009003.
Spasojevic, S., Liu, L. & Gurnis, M., 2009. Adjoint models of mantle convection with seismic, plate motion, and stratigraphic constraints: North America since the late cretaceous, *Geochem. Geophys. Geosyst.,* **10**(5), Q05W02, doi:10.1029/2008GC002345.
Stadler, G., Gurnis, M., Burstedde, C., Wilcox, L.C., Alisic, L. & Ghattas, O., 2010. The dynamics of plate tectonics and mantle flow: from local to global scales, *Science,* **329**(5995), 1033–1038.
Stocker, R.L. & Ashby, M.F., 1973. On the rheology of the upper mantle, *Reviews of Geophysics,* **11**(2), 391–426.
Tarantola, A., 2005. *Inverse Problem Theory and Methods for Model Parameter Estimation,* SIAM.
Tröltzsch, F., 2010. *Optimal Control of Partial Differential Equations: Theory, Methods and Applications,* Vol. 112 of Graduate Studies in Mathematics, American Mathematical Society.
Vogel, C.R., 2002. *Computational Methods for Inverse Problems,* Vol. 23, SIAM.
Wells, R.E. & McCaffrey, W., 2013. Steady rotation of the Cascade arc, *Geology,* **41**(9), 127–130.
Worthen, J., Stadler, G., Petra, N., Gurnis, M. & Ghattas, O., 2014. Towards an adjoint-based inversion for rheological parameters in nonlinear viscous mantle flow, *Phys. Earth planet. Int.,* **234,** 23–34.
Zhong, S., Gurnis, M. & Moresi, L., 1998. Role of faults, nonlinear rheology, and viscosity structure in generating plates from instantaneous mantle flow models, *J. geophys. Res.: Solid Earth,* **103**(B7), 15 255–15 268.

APPENDIX A: INTEGRATION OF VISCOSITY BOUNDS IN RHEOLOGY

The classical Newton method requires continuous differentiability of the non-linear equation it is applied to. Thus, we incorporate the viscosity bounds into the strain-rate weakening viscosity such that the map from the second invariant of the strain rate (and thus from the velocity) to the stress tensor is differentiable. We do so by choosing the shift d in (5) appropriately. The strain rate $\hat{\dot{\varepsilon}}_{\mathrm{II}}$, where the upper viscosity bound $\eta_{\max}$ becomes active is characterized by

$$\eta_{\max}\dot{\varepsilon}(u) = a(T)(\dot{\varepsilon}_{\mathrm{II}} - d)^{\frac{1}{2n}}\dot{\varepsilon}_{\mathrm{II}}^{-\frac{1}{2}}\dot{\varepsilon}(u),$$

which implies that

$$\eta_{\max}\hat{\dot{\varepsilon}}_{\mathrm{II}}^{\frac{1}{2}} = a(T)(\hat{\dot{\varepsilon}}_{\mathrm{II}} - d)^{\frac{1}{2n}}. \tag{A1}$$

Solving for the shift d, this results in

$$d = \hat{\dot{\varepsilon}}_{\mathrm{II}} - \left(\frac{\eta_{\max}}{a(T)}\right)^{2n}\hat{\dot{\varepsilon}}_{\mathrm{II}}^{n}. \tag{A2}$$

To ensure differentiability of the stress tensor with respect to the second invariant of the strain rate, we will choose the shift d such that the first derivatives of both sides in (A1) coincide at $\hat{\dot{\varepsilon}}_{\mathrm{II}}$, that is,

$$\frac{1}{2}\eta_{\max}\hat{\dot{\varepsilon}}_{\mathrm{II}}^{-\frac{1}{2}} = a(T)\frac{1}{2n}(\hat{\dot{\varepsilon}}_{\mathrm{II}} - d)^{\frac{1-2n}{2n}}.$$

Using (A1), this implies that

$$\hat{\dot{\varepsilon}}_{\mathrm{II}}^{-1} = \frac{1}{n}(\hat{\dot{\varepsilon}}_{\mathrm{II}} - d)^{-1},$$

and thus necessarily that $\hat{\dot{\varepsilon}}_{\mathrm{II}} \geq d$. Hence,

$$d = \frac{n-1}{n}\hat{\dot{\varepsilon}}_{\mathrm{II}}. \tag{A3}$$

Substituting d from (A2) into (A3) yields

$$\left(\frac{\eta_{\max}}{a(T)}\right)^{2n} \hat{\dot{\varepsilon}}_{\mathrm{II}}^{n} = \frac{1}{n}\hat{\dot{\varepsilon}}_{\mathrm{II}},$$

resulting in

$$\hat{\dot{\varepsilon}}_{\mathrm{II}} = n^{\frac{1}{1-n}}\left(\frac{\eta_{\max}}{a(T)}\right)^{\frac{2n}{1-n}} \tag{A4}$$

Substituting $\hat{\dot{\varepsilon}}_{\mathrm{II}}$ into (A2) gives the desired expression for d, which is independent of the strain rate:

$$d = (n-1)n^{\frac{n}{1-n}}\left(\frac{\eta_{\max}}{a(T)}\right)^{\frac{2n}{1-n}}.$$

Note that when the rheology has a strain rate exponent of $n = 1$, the shift d is not well defined. However, in this case, the rheology is Newtonian and the viscosity is independent of $\dot{\varepsilon}_{\mathrm{II}}$, and the effective viscosity (5) simplifies.